

SYNTACTIC THEORY
AND
THE GRAMMAR OF CATALAN COMPOUNDS

Anna Gavarró Alguero

PhD
University of Edinburgh

1990



Declaration

I declare that this thesis has been composed by myself and that the research therein has been conducted by myself unless otherwise indicated. Material included in Chapters One, Three and Four has appeared as Gavarró 1988, Gavarró 1989 and Gavarró to appear, in *Work in Progress*, Department of Linguistics, University of Edinburgh.

Anna Gavarró Alguero

Edinburgh, October 15th, 1990.

Acknowledgements

During my stay at the University of Edinburgh, I have received an Exchange Scholarship from the Stevenson Foundation for the academic year 1987-88, a scholarship from the CIRIT of the Generalitat de Catalunya for the summer of 1988, and two studentships from the Vans Dunlop Foundation of the Faculty of Arts of the University of Edinburgh, during the academic years of 1988-89 and 1989-90. I am grateful to them for having, together with my family, financed me for these four years.

In the Department of Linguistics, I am particularly indebted to my supervisor, Ronnie Cann, for having discussed with me many issues, and having provided much advice and encouragement, and I am grateful to Jim Miller, for his valuable comments, and help. Also, I would like to acknowledge Irene MacLeod and Morag Brown for their technical assistance. There are many other people to whom I owe thanks, especially in the Department of Linguistics and the Centre for Cognitive Science; among them, I would like to mention Bruce Connell, Vernon Shilliday, and Elina Rigler. Special thanks to Glyn Morrill.

My parents, Ròmul Gavarró and Montserrat Algueró, have given me their unconditional support. I dedicate this thesis to them, as to the memory of my grandparents Josep Algueró and Raimunda Pagès.

Abstract

According to a prevalent lexicalist view, morphology is autonomous from syntax: syntactic principles do not apply to morphological structures. This thesis develops an alternative view, represented in the work of Baker (1988a). In this approach, some morphological phenomena are treated in a Principles and Parameters model of syntax (cf. Chomsky 1981, 1986a), augmented with a subtheory of morphology. This kind of word formation is therefore hypothesised to be subject to the Projection Principle and the principles of the current subtheories of syntax, in addition to the morphological subtheory, which constrains the possible X^0 structures. The resulting framework is applied to compounding in Catalan.

The account postulated provides a model which is restrictive both with respect to the constraints of the grammar and to the information contained in lexical entries. The logically possible compound forms are systematically surveyed; grammaticality is determined by reference, in particular, to the interaction of theta theory, Case theory, and government theory, together with the morphological subtheory. However, no reference is made to Move α , unlike in Baker's treatment of incorporation and other syntactic accounts of word formation. [N V] verbal compounds, whose interpretation involves inalienable possession, and which are apparently unique to Catalan amongst the Romance languages, raise the question as to how to formalise inalienability; this notion is shown to be derivable if morphological structures are constrained by binding theory. Parameterisation of morphological principles is considered; there is discussion of the significance for universal grammar of the taxonomy of compounds and compound-like words induced by the role of Case theory in the analysis.

Abbreviations

A(P)	adjective (phrase)
ABL	ablative
ACC	accusative
AG	agent
Adv	adverb
AOR	aorist
BEN	benefactive
cl	clitic
CONJ	conjunction
DAT	dative
DET	determiner
DIM	diminutive
EXP	experiencer
FEM	feminine
IND	indicative
INF	infinitive
INSTR	instrumental
LOC	locative
N(P)	noun (phrase)
NEUT	neuter
NOM	nominative
num	numeral
O	object
onom	onomatopoeia
P(P)	preposition (phrase)
PL	plural
pp	past participle
prp	present participle
pr	present
REFL	reflexive
S	sentence
SG	singular
TH	theme
V(P)	verb (phrase)

Table of Contents

<i>Chapter One. Introduction</i>	1
1.1. Data	3
1.2. Semantics	4
1.3. On idiosyncrasy	5
 <i>Chapter Two. Data and review of the literature</i>	 13
2.1. Fabra	14
2.2. Badia	19
2.3. Moll	22
2.4. Ferrater	24
2.5. Wheeler	26
2.6. Grossmann	30
2.7. Cabré and Rigau	33
2.8. Mascaró	36
2.9. Duarte and Alsina	42
 <i>Chapter Three. Compounding in GB; theories of meaning</i>	 44
3.1. Framework	44
3.1.1. Lexical entries	45
3.1.2. Principles of the grammar	49
3.1.2.1. Morphological versus lexical	50
3.1.2.2. GB and word formation	51
3.2. On some universal constraints in morphological theory	52
3.3. Subtheories of meaning	54
3.3.1. Basic assumptions	54
3.3.1.1. Theta theory	62
3.3.1.2. Predication theory	63
3.3.2. Compounding and theta theory	69
3.3.2.1. The <i>alatrencar</i> type	69
3.3.2.1.1. On inalienability	77
3.3.2.1.2. Anaphoric islands	82
3.3.2.2. The <i>parallamps</i> type	86
 <i>Chapter Four. Case theory</i>	 95
4.1. A generalised theory of Case	95
4.2. Compounding and Case theory	99

4.2.1. Why not Move α ?	101
4.3. Catalan compounds and Case	106
4.4. Case theory and X-bar theory	109
4.4.1. X-bar theory and word formation	111
4.4.2. Feature Percolation Conventions	113
4.4.3. Some speculations on word order	122
<i>Chapter Five. Government theory and structural conditions</i>	127
5.1. Government theory and categorial constraints	127
5.1.1. Government and word formation	127
5.1.1.1. On the notion of government	127
5.1.1.2. A problem	130
5.1.1.3. Government theory and compounding	132
5.1.2. Categorial constraints	135
5.1.2.1. The adverbs in <i>-ment</i>	135
5.1.2.2. Extending the analysis	140
5.1.2.3. INFL	142
5.1.2.4. Some other issues of syntactic category	144
5.2. Universal constraints in morphology revisited	147
<i>Chapter Six. Catalan compounds</i>	151
6.1. The right-headed compounds of Catalan	151
6.1.1. [N N]	153
6.1.2. [V N]	155
6.1.3. [A N]	156
6.1.4. [P N]	157
6.1.5. [N V]	159
6.1.6. [V V]	163
6.1.7. [A V]	163
6.1.8. [P V]	166
6.1.9. [N A]	168
6.1.10. [V A]	173
6.1.11. [A A]	173
6.1.12. [P A]	175
6.1.13. [X P]	175
6.2. Other semantic relations	176
6.3. A remaining issue: The case of <i>enriure's</i>	180
6.4. Lexicalised expressions	183
6.5. Catalan compounding in the context of Romance	191

<i>Chapter Seven. Conclusion</i>	194
7.1. Morphological theory	194
7.1.1. The Projection Principle	196
7.1.2. Some principles of morphological theory	197
7.1.3. The characterisation of compounds	199
7.2. Topics for further research	200
 <i>References</i>	 202

Chapter One

Introduction

The purpose of this thesis is twofold: on the one hand I intend to give a comprehensive account of compounding in Catalan; on the other hand, I explore the possibility of a syntactic rendering of compounding, a phenomenon found across the languages of the world.

Compounding has often been considered to be the most syntactic of all word formation phenomena (leaving inflection aside). It makes no sense, though, to loosely consider a linguistic phenomenon more or less syntactic, if syntacticity is a discrete notion: a phenomenon is either construed at the levels of representation known as syntax or not, and accordingly it is syntactic or it is not. I want to substantiate the view according to which compounds conform to syntactic principles. The framework chosen here (and within which syntactic principles are defined) is a Principles and Parameters grammar along the lines of Chomsky (1981, 1986a).

The basic apparatus of such a linguistic model consists of a set of primitive elements and a set of principles and rules of grammar which determine which combinations of the primitive elements are possible and which are not. These primitive elements (or lexical items) are characterised in lexical entries. The principles are grouped in various subtheories (e.g. theta theory, Case theory, binding theory). The only rule belonging to this grammar is a rule of movement of constituents, Move α . Both rules and principles apply freely whenever their structural descriptions are met, and the basic structures to which these apply result from the free concatenation of lexical items.

The relations holding in the syntactic representation of structures cannot be altered without limits, given the Projection Principle, which ensures that the basic features of lexical items remain unchanged:

- (1) Representations at each syntactic level (i.e., LF, and D- and S- structure) are projected from the lexicon, in that they observe the subcategorisation properties of lexical items.
(Chomsky 1981:29)

If a grammar is thought of as described above, the characterisation of compounds constitutes a particular problem, namely the establishment of well-formedness conditions for $[X^0 Y^0]_{Z^0}$ structures, where X^0 and Y^0 are not affixal. (This widely accepted definition of compound will be further refined in the course of this thesis, mainly by reference to Case theory).

In anticipation of the approach, it is worth pointing out a potential ambiguity of the term "lexical". Because compounds are words (and not phrases or sentences), a study of compounds is a lexical study -- where this expression is meant to refer to the study of words. This is the first meaning of the term lexical. Second, lexical is often used interchangeably with "belonging to the lexicon". If compounding takes place in the syntax on the basis of the information drawn from the lexicon, then compounding is not lexical in the second sense of the word. The two readings of "lexical" can only be taken to go together if the Strong Lexicalist Hypothesis (cf. Di Sciullo and Williams 1987) is maintained, since this identifies "word formation" and "formation in the lexicon". This hypothesis will not be maintained here.

The structure of this thesis is as follows. Chapter Two presents data on Catalan compounds collected by several linguists, and summarises their analyses and comments. Chapters Three, Four and Five present a Government and Binding approach to compounding, and some detailed analyses of particular problems illustrated by Catalan compounds. Unlike usual syntactic treatments of word building, the one proposed here does not involve Move α , and so is non-transformational. Chapter Three introduces Government and Binding theory; it also discusses issues related to meaning in this framework, and argues for the relevance of theta theory and binding theory in word formation. Chapter Four is concerned with the way in which meaning relations materialise, and thus involves Case theory, and reference to X-bar syntax. In Chapter Five I consider the application of government theory to compounding, as well as some categorial constraints which hold on $[X^0 Y^0]_{Z^0}$ structures. Chapter Six presents systematically the logically possible combinations of lexical items at the base of compounds and discusses the predictions made, for Catalan, by the grammar outlined. The dissertation concludes in Chapter Seven with a general overview of the morphological theory that this analysis of compounds is conceived as a part of. Finally, some topics for future research are suggested. In the remainder of this chapter, I briefly mention the methodological approach to data engaged here, and indicate the kind of semantic specifications I occasionally use in the thesis. I also

address a general issue, namely idiosyncrasy, which bears on my approach to lexical (i.e. word-) studies.

Throughout, examples in languages other than English will be glossed, and translations will be supplied when the gloss is not sufficient. When they are not already separated by blank spaces in standard orthography, Catalan complex expressions will have their constituents separated by hyphens. Catalan orthography makes use of hyphens as well, which means that my notation blurs some orthographic distinctions (e.g. *pit-roig* is written with a hyphen, but *parallamps* is not; yet both appear with a hyphen in my examples). Anyway, orthography is irrelevant for our concerns and the purpose of my notation is to homogenise the transcription of complex words regardless of their orthography. Abbreviations found in the glosses as well as those used in the text are listed before the table of contents.

Finally, the following convention is adopted: "Baker 1989", e.g., refers to a piece of work, while "Baker (1989)" is a reference to the author mentioned in connection with work published in 1989.

1.1. Data

Despite the fact that Chapter Two assembles the data on Catalan compounds presented by many authors in the recent past, I aim at accounting for my intuitions as a native speaker of Catalan. These often coincide with those presented in other works; when they do not, this is pointed out.

Unlike those who claim that lexical (i.e. word-) studies presuppose the collection of a large range of data found in dictionaries or provided by several speakers, I subscribe to the view expressed in quite radical terms by Corbin (1987:46) -- which has come to my knowledge through Zwanenburg's (1989) review -- according to whom, first, the linguist's object

"is larger than everything which one finds in all dictionaries, in that its sole limits are linguistic in nature; and it is more limited than what one finds in the dictionaries, in that it must be stripped of everything which is not strictly linguistic. This double perspective, of *reconstruction* and *purification* of the object to be described in relation to the lexicographer's material, constitutes the originality of the theory which I propose, as to the definition of its object".

Corbin (1987:81-82) continues:

"it is better to consult one metalinguistically non-naïve speaker than several unsophisticated speakers"

given that language users tend to be unable to separate their lexical competence

and their conventional lexical knowledge. In short, the linguist himself or herself is a natural candidate as metalinguistically non-naïve native speaker of his or her own language.

Though Corbin might not want to maintain her methodological stand for non-lexical judgements (i.e. judgements of sentence grammaticality), her methodology has a desirable effect when applied to sentence judgements too: it guarantees that heterogeneous data (psychologically heterogeneous data, thus heterogeneous from the point of view of the theory adopted here) are avoided as much as possible. Collections of data from several speakers may not be consistent, that is, consistently generated by any (psychologically real) grammar.

Putting aside the problems of retrieving one's intuitions (cf. Bourdieu et al. 1968 on the problem of the construction of data), it seems reasonable to assume that a single speaker's intuitions, such as one's own, reflect a single psychological grammar. Note that the notion of language as social construct plays no role in the present theory. Whenever the expression "the grammar of Catalan" is used, it must be understood that claims are only being made about one or several speakers of Catalan, not necessarily all of them, given that not all speakers of Catalan share exactly the same internalised grammar. Individual speakers' intuitions provide information as to individual points between which language variation exists, and in this way single speaker intuitions underpin the ultimate goals of Chomskyan linguistics.

1.2. Semantics

On a few occasions I have chosen one analysis amongst several on the basis of semantics alone. Yet, the issue of the semantic interpretation of the syntactic representations postulated in GB grammar remains unresolved. Meaning relations are encoded, in a GB grammar, at D-structure, and most importantly at LF. Even if we agree that some problems which semantic theories address are handled by GB syntax (cf. Newmeyer 1986), it seems that a way should be found of systematically relating a GB analysis with a model theoretic interpretation.

It will be assumed that LF -- where the scope of quantifiers, etc., is encoded -- corresponds most closely to a formal semantic representation of meaning. Model theoretic interpretation is based on analyses of meanings of expressions into functions and arguments (cf. Allwood et al. 1977 for an introduction to model theoretic semantics, and Dowty et al. 1981 for a Montague semantics approach in particular). When considered necessary, the meanings of expressions will be represented here by terms of the lambda calculus. These are built up from

constants and variables by means of *application* and *abstraction*. If α and β are terms, then the application of α to β ($\alpha \beta$) represents the result of applying the function α to the argument β . If α is a term and x is a variable, then the abstraction of α over x , $(\lambda x \alpha)$ represents the function which when applied to an argument β gives $\alpha[\beta/x]$, i.e. the result of substituting β for the free occurrences of x in α . Hence the fundamental law of the lambda calculus, based on these definitions, known as lambda conversion:

$$(2) ((\lambda x \alpha) \beta) = \alpha[\beta/x]$$

Constants will be represented in boldface and outermost parenthesis will be dropped in all representations.

1.3. On idiosyncrasy

Addressing the issue of idiosyncrasy seems inevitable when discussing any kind of word formation. In this section, I intend to show that the link between idiosyncrasy and words is less essential than is often supposed. In order to do so, I shall first discuss the notion of idiosyncrasy in quite general terms (granted that linguistic idiosyncrasy is a particular case of idiosyncrasy in the colloquial sense of the term). Second, I shall consider more specific implications that this notion has in the construction of a grammar. Most important, the concept of idiosyncrasy excludes a kind of analysis which is common in the linguistic literature, and yet misleading, namely an analysis implying that there is such a thing as predictable idiosyncrasy.

The examples in (3) illustrate what is meant by idiosyncrasy. It is clear that part of the meaning of these compounds of Catalan cannot be predicted from the meaning of their constituents; i.e. that even a native speaker would not be able, when hearing them for the first time, out of context, to ascertain what it is that they designate.

- | | | |
|-----|----------------|--------------------------|
| (3) | mata-parents | |
| | kill relatives | (kind of mushroom) |
| | perdona-vides | |
| | forgive lives | 'bully' |
| | espia-dimonis | |
| | spy demons | 'dragon-fly' |
| | toca-campanes | |
| | ring bells | 'feather-brained person' |

In some cases, in fact, the only semantic information that the compound conveys is that it is nominal, verbal, or whatever is the case; i.e. it only points to the semantic type(s) which can be associated with its syntactic category. This is exemplified by

mataparents, which can be identified as a noun, but where the connection between what is designated and the literal meaning of the compound is not identifiable. There are other instances of idiosyncrasy, such as words which are phonologically odd, or which display irregular morphology. These are problems, related to allomorphy and so on, which I am not going to deal with here.

However, the domain of idiosyncrasy is not only the word; there are phrases whose meaning is not a function of the meaning of their parts, that is, whose meaning is not compositional¹. This is what is usually called an idiom -- though I suggest below that the use of this term might be extended to constructions which traditionally have not been called idioms. (4) exemplifies some traditionally recognised idioms.

- | | | |
|-----|--------------------|---------------------|
| (4) | somniar truites | |
| | dream-INF trouts | 'to be a dreamer' |
| | tocar campanes | |
| | play-INF bells | 'to miss the point' |
| | fer denteta | |
| | make-INF tooth-DIM | 'to tantalise' |
| | anar de tort | |
| | go-INF of bent | 'to stagger along' |

These belong to the class of idiomatic expressions as tentatively defined by Bar-Hillel (1954:192):

"An expression in a given language L is idiomatic within L, with respect to a given monolingual dictionary and a given list of grammatical rules if, and only if, none of the word sequences correlated to the given expressions by the dictionary and the list of rules is (sufficiently) synonymous with it."

Regardless of the kind of grammar postulated (with rules or, as in GB, with well-formedness conditions plus a rule like Move α), idiomatic expressions are expressions whose meaning is unpredictable. How, then, can this unpredictable information be included in the grammar?

It seems, insofar as idiomatic expressions are idiosyncratic, that the only thing that can be done is to list them. If they cannot be reduced to any kind of regularity, generalisations are impossible. idiosyncrasy is, by definition, a characteristic of individuals, of units and not classes. Once it has been established that idiomatic expressions (indeed all manifestations of idiosyncrasy) must be listed, the question arises as to where to list them in a grammar.

¹Or, rather, they can only be interpreted compositionally if one states somewhere (possibly in the lexicon) that the composition of one element with the other gives the (unpredictable) meaning it has. But, even so, the idiosyncrasy of this meaning is expressed by the lack of generality of the rule of composition required.

The lexicon has often been conceived as the locus of idiosyncrasies, that is of all the information which is not predicted by the productive processes of the grammar. The simple words in the lexicon are naturally considered idiosyncratic, thus the fact that *gos* designates 'dog' and *gat* 'cat', in Catalan, is arbitrary from a completely synchronic point of view; we could have a language where *gat* designated 'dog' and *gos* 'cat' and the structure of the two languages would not be altered. (This is nothing more than Saussure's (1916) arbitrariness of signs).

If the lexicon is thought of as the collection of idiosyncrasies, clearly it must contain entries for idioms, such as those in (4), as well as simple words. In current practice it is uncontroversial to assume that such expressions belong to the lexicon. Lexical entries, therefore, are not entries for words (or words and affixes) exclusively. There are entries for idiosyncratic phrases too. A grammar which places words (and affixes) in the lexicon and all phrases outside is empirically inadequate to handle idioms.

So far I have made the point that idiosyncrasy cannot be associated only with words. But also, it must be remarked that regularity is not exclusive to expressions larger than the word (i.e. phrases and sentences). In fact, this is the idea behind any study of word formation, generative or otherwise. There is no a priori reason to think that idiosyncrasy is a characteristic of all words (see for instance Baker 1988a, for highly regular word formation).

The question then is: Where should the regularity of word formation be expressed? It has been assumed for a long time that its place is the lexicon, and this is the position taken by the followers of the Strong Lexicalist Hypothesis; others defend the position according to which at least certain instances of word formation can take place in the syntax (i.e. at levels of representation primarily postulated for sentence formation, D-structure, S-structure, etc.). As a consequence, at least in a model subscribing to the Strong Lexicalist Hypothesis, the lexicon cannot be considered any longer the locus of idiosyncrasy only².

Briefly, the alliances word/idiosyncrasy and phrase/regularity cannot be accepted if we want our grammar to reach even observational adequacy. The link word/idiosyncrasy is only licit as a generalisation when referring to simple words. These are certainly idiosyncratic, i.e. singular, and so need individual specification. (The identification of word and idiosyncrasy, so typical of the first years of research in generative grammar -- and which had as consequence the lack of interest for word formation, cf. Scalise 1984 -- might originate in the inherent

²Botha (1968) makes the point that the notion of a lexicon conceived as list of idiosyncrasies is made inconsistent if any predictable information is included in it (and yet partial regularities should not be lost, but included in the grammar). This is perfectly compatible with my approach.

idiosyncrasy of simple words). Simple words constitute the vocabulary of Bar-Hillel's dictionary, the set of primitives which are at the base of any linguistic structure. As such, simple words are grammatical in virtue of being listed in the grammar, rather than in virtue of being generated.

A terminological clarification is then in order: idiosyncratic is a qualification which applies to simple words and some complex words and phrases. Idiomatic expressions are those whose meaning is unpredictable (i.e. idiosyncratic in a particular fashion) and are structurally complex. According to this, then, there is nothing strange in considering some compounds to be idioms.

To summarise, idiosyncratic forms are listed, and the lexicon seems the ideal place for that. On the other hand, whether some kinds of regularity are expressed inside the lexicon as well, or outside it, is controversial³. Simple words form a particular class inside the lexicon -- the class which Bar-Hillel calls dictionary. Note that, if we identified our lexicon with his dictionary, it would follow that there would be no idiomatic expressions at all.

More crucial than *where* to locate idiosyncrasy is the question of *what* can be considered idiosyncratic. Bar-Hillel's definition emphasises the dialectic between rules (or principles) and idioms, once a dictionary has been established. Probably no linguist would claim that everything in a natural language is reducible to regularity, i.e. that rules and principles apply without exception. The empirical problem remains of deciding what is an exception and what a rule or a principle which has been overridden. This is indeed one of the basic tasks of the linguist. The true testing ground for the interaction of regularity and idiosyncrasy is related to productivity. I shall consider a form regular when a native speaker can create and/or understand new expressions of the same kind. Only if a new expression conforms to some regularity can it be predicted to be possible⁴. As for what exists, it may be regular or not, idiosyncratic or not.

In particular, one might want to consider that the mere existence of a complex expression is contingent, so that its existence is one of the expression's idiosyncrasies (the only one, if the form is regular in all respects). To put it another way: a complex word -- formed according to the principles of the language -- may exist or be only potential for a speaker; the acknowledgement of its existence is, thus, something which the speaker must associate with that word only -- its

³Pesetsky (1985) bears on this issue, and claims that a lexicon conceived as a storehouse of idiosyncratic properties of lexical items is not a coherent *component* of the grammar, understanding by component a level of representation with well-defined properties. His point is that the idiosyncrasies stored become operative at different levels of representation.

⁴By this I am not entirely precluding the possibility of something being both regular and unproductive.

existence is something peculiar to it, therefore idiosyncratic. This assertion is consistent with a position maintained in Jackendoff 1975, which says roughly that speakers are aware of the words they have heard, so that their existence must be encoded in the grammar (though this might be too strong for inflected forms, it seems to be quite true of e.g. compounding in Romance; Baker (1988a) and Mithun (1984) claim it is also true of incorporation). If this is so, this clearly differentiates part of word formation from sentence formation: the output of word formation is recorded by the speaker/hearer, unlike the output of sentence formation. (In contrast, the fact that a lot of sentences are memorised by the speakers of a language is attributed to their encyclopaedic knowledge, not to any linguistic factor). It follows from this approach that the output of word formation is always idiosyncratic in at least one respect: words are marked as existent or not. If so, we have to slightly modify the picture given above; complex words are not necessarily idiosyncratic, *except* with regard to the unpredictable factor of their existence.

I have shown that idiosyncrasy does not relate only to words, but also to phrases. Second, by definition, idiosyncrasy and prediction are mutually exclusive. Third, it is assumed that complex words can be completely regular, though their existence cannot be predicted, and thus must be indicated in the grammar (constituting, therefore, an idiosyncratic feature).

Together with (4), (5) illustrates the idiomatic expressions found in Catalan.

(5)	a-les-hores in the hours	'then'	*
	tot-hora all hour	'always'	
	a la vora to the edge	'near'	
	qual-se-vol which SE want	'any'	
	sobre-tot over all	'especially'	
	per-tot for all	'everywhere'	
	tan-mateix so same	'nevertheless'	
	tam-bé so good	'also'	
	en-l-aire in the air	'up'	

en-lloc	
in place	'nowhere'
a-l-hora	
in the hour	'at the same time'

The patterns (pairs of structure and meaning) to which the idiomatic expressions in (5) conform do not give rise to well-formed new compounds in Catalan. Later, I shall discuss the reason why these patterns are not productive (so that, for example, **cap-hora* 'no-hour' meaning 'never' is ungrammatical, while *tothora* is good). Assuming this, however, I consider it legitimate to give no account of the expressions in (5), because, insofar as they are idiomatic, the grammar cannot predict them. A descriptively adequate grammar accounts for them by listing them, not only as existent words, but as idioms -- with specific meaning and structure. The question above, of the reason why newly created compounds of this sort would be ungrammatical, requires a more explanatory answer.

The position held here is that whether an expression is idiosyncratic or not is not uncontroversial, but once the grammarian considers it idiosyncratic, s/he gives up hope of accounting for it. Linguistically idiosyncratic, then, means irreducible to a linguistic account.

A paradox seems to arise with expressions such as those in (5), exemplified here by *aleshores* 'then'. This adverb is peculiar in at least one respect: its meaning is unpredictable: 'in the hours' meaning 'then'. Second, arguably, it is syntactically idiosyncratic if it is an X^0 projection instead of a maximal projection of form $[P\ NP]_{pp}$. X^0 projections (as we shall see) do not display word-internal agreement such as that exhibited between *les* and *hores*, both $[+PL]$ and $[+FEM]$.

Given this state of affairs, the following paradox arises: *aleshores* is: (i) grammatical because it belongs to the lexicon, and this is included in the grammar; (ii) predicted ill-formed with its actual interpretation, given the regular mechanisms of semantic interpretation (and possibly syntactically ill-formed as well). So, by putting (i) and (ii) together, we find ourselves in the paradoxical situation of finding *aleshores* both grammatical and ungrammatical (and naturally these two notions are mutually exclusive).

Two ways out of this apparent contradiction suggest themselves. In the first, *aleshores* is grammatical because the grammar does not analyse it, but rather takes it as a unit. So *aleshores* has a lexical entry in which it is identified as an adverb, but there is no indication of it being a complex word.

The alternative solution makes *aleshores* grammatical because of it being an existent word, listed in the lexicon. The word's existence overrides all grammaticality predictions made by (the generative component of) the grammar.

This solution, unlike the first one, does not exclude the analysis of *aleshores* as [P NP], and this analysis is available to the speakers of the language. So, for words which speakers can analyse as complex, but whose structure is unproductive, the second solution is more appropriate. The fact that those words have internal structure may be empirically tested by looking, for example, to their phonological form: if the components of a complex word are still perceived as words, they preserve some phonological characteristics of words (vowel quality, etc.). Phonological erosion is, on the other hand, the consequence of structural (and semantic) opacity. (In this connection, see Cutler 1982).

The first solution might be of some use in considering the historical evolution of words. Initially, complex words are analysed by the speaker. Later, it is possible (though not necessary) that the lack of productivity of their structures facilitates their opacity. They are no longer analysed by speakers, and so the lexicon includes them as units. Or, as summarised by Darmesteter (1875:XI):

"L'existence des mots composés comprend deux époques distinctes: celle où ils apparaissent comme composés et celle où ils deviennent simples".

In intermediate stages, speakers may be able to find analyses for the word, but do not normally assign one to it. (Notice, nevertheless, that this historical process remains unexplained).

I have argued so far for the necessity of handling idiosyncrasy in the grammar, whether it is limited to a word or not. In this approach, idiosyncrasy cannot be linked to any component⁵ of the grammar, nor to specific principles.

This line of thought contrasts with that of other linguists, who appeal to idiosyncrasy to account for different kinds of word formation. As an example among many others, Borer (1988) argues that the nominal compounds of Hebrew are formed in the lexicon and *therefore* are often semantically opaque. On the other hand, the Hebrew construct state nominals are semantically transparent and thus formed in the syntax. Her reasoning is based on the assumption that the formation of Hebrew compounds is describable and, at the same time, gives rise to idiosyncrasies (i.e. does not conform to a regular pattern). Briefly, her position can only be maintained if we are prepared to accept that idiosyncrasy is predictable. However, such a statement is contradictory, and an analysis based on it is unsound. Independently of whether Hebrew compounds should be formed in the lexicon or in the syntax (I leave aside other arguments put forward by Borer in her analysis), it should be clear that the argument involving idiosyncrasy is deceptive.

In conclusion, I have argued that an idiosyncrasy is to be understood by its

⁵See footnote 4.

definition to be a property of a singular item, so that reference to it affords no generality, and can only be a description of fact, never an explanation.

Chapter Two

Data and review of the literature

This chapter describes the work on Catalan compounds by (as far as I know) all the authors whose work has been published on this subject during the last 35 years; I leave aside all grammars developed for educational purposes, etc., with less emphasis on theoretical matters. I intend to give in some detail the data they have looked at, as well as offer a critical presentation of their work. Although all the data will be treated in subsequent chapters, they will not all be construed as compounding in the technical sense defined on the basis of that treatment.

I shall examine the work of: Pompeu Fabra, Antoni M. Badia i Margarit, Francesc de B. Moll, Gabriel Ferrater, Max W. Wheeler, Maria Grossmann, Teresa Cabré and Gemma Rigau, Joan Mascaró, and Alex Alsina and Carles Duarte. Their respective works are not all, as one could expect, of the same value and/or influence, nor are they all conceived as belonging to the same paradigm, and this will be easily perceivable in the review. Fabra's work can be considered an early structuralist effort to characterise Catalan. Also as structuralists I would classify Badia i Margarit, Moll and Grossmann, this last one in the tradition of Eugenio Coseriu. As such, they are all interested in languages (as opposed to grammars -- in Chomskyan terminology); I shall try to assess these works in their own terms. Ferrater's work on compounds was written under the guide of Emile Benveniste's (1974a, 1974b) classification, though Ferrater was the first to be concerned with the construction of a grammar in the Chomskyan sense. The approaches of Wheeler, Mascaró, Cabré and Rigau share this feature, and are circumscribed in the transformational generative grammar framework (which kind of transformational generative grammar is highly dependent on the chronology). Duarte and Alsina's work on compounds is almost exclusively taxonomic.

2.1. Fabra

Pompeu Fabra is certainly the most influential of all Catalan linguists I shall mention; he is referred to in most subsequent studies and indeed his work is the basis for many of them. (He was the author of both the normativisation of modern Catalan and its accepted orthography). In his posthumous *Gramàtica catalana* [*Catalan Grammar*], published in 1956, he devotes a chapter to compounding.

Fabra groups compounds into two main classes: "composts formats amb prefixos" ('compounds formed with prefixes'), e.g. (1a-d), and "els altres composts" ('other compounds'), e.g. (1e-h)¹:

- (1) a. em-bagular
in trunk-INF 'to put into a trunk'
 or 'to fill with trunks'
- restablir
 re establish-INF
- b. entre-cella (space between the eyebrows)
 between brow
- inter-planetari
 inter planetary
- c. ante-cedent
 ante-cedent
- d. neo-classicisme
 neo classicism
- e. compta-gotes
 count drops 'dropper'
- f. ala-llarg
 wing long 'long winged'
- g. lliri-jonc
 lily rush 'sword lily'
- h. blat de moro
 wheat of Moor 'maize'

These two classes are further divided. Of the "compounds formed with prefixes", (1a) exemplifies non-stressed prefixes, (1b) stressed prefixes, (1c) prefixes which were taken from Latin and (1d) those taken from Greek. (1b) includes late Latin borrowings, such as *inter-* in *interplanetari*, too, and (1c) contains only words for

¹In the glosses, I translate e.g. *compta* in (1e) as 'count' because it can correspond to the second person singular imperative or to a basic form of the verb; it could also be glossed as 'counts' since it coincides, too, with the third singular of the present indicative (in fact, this is the gloss chosen in Wheeler 1979). This issue is discussed further in what follows.

which segmentation into Catalan words is not possible; Fabra himself says that they *were* Latin complex words: consider *antecedent* 'antecedent', *incògnit* 'unknown', *inaudit* 'unheard-of', for which the respective *cedent*, *cògnit* and *audit* are nonexistent in Catalan. Altogether, (1a-d) lists Latinate and Greek vocabulary, together with what Fabra calls prefixed words. On this point, Fabra seems to fall into an inconsistency: "compounds formed with prefixes" is a contradiction in terms, given his own definitions of compounding and derivation, which includes prefixation. He has characterised compounds this way (Fabra 1956:142):

"en aquells mots ambdós elements constitutius són mots" ('in those words both constituent elements are words')

to add later:

"Quan el primer element constitutiu d'un mot compost és un adverbí o una preposició, aquest element és considerat com un prefix, i així pot dir-se que un mot com *contraprojecte* és un derivat de *projecte* per mitjà del prefix *contra*, com es diu que *ventós* és un derivat de *vent* per mitjà del sufix *ós*." ('When the first constituent of a compound is an adverb or a preposition, this element is considered like a prefix, and so one can say that a word like *contraprojecte* 'counterproject' is a derivative of *projecte* 'project' by means of the prefix *contra* 'against', as one says that *ventós* 'windy' is a derivative of *vent* 'wind' by means of the suffix *-ós* '-y').

The characterisation of derivatives (Fabra 1956:143) is incompatible with the one of compounds:

"en aquells mots un sol dels seus components és un mot" ('in these words, only one of the elements is a word').

Thus stressed and non-stressed prepositions and adverbs lose their status of words when occurring before another word, which has a prefix and, therefore, is derived (as opposed to compound). Although inconsistent, the approach of Fabra on prefixes and compounds is followed by many others, even those aware of its deficiencies. (For an exposition of these, see Cabré 1988.)

Let us look at two cases in particular: that of *ben-*, *mal-* and that of *en-*. *Ben* and *mal* are considered prefixes e.g. in *maldestre* (from *mal* 'badly' and *destre* 'skilful') and *malfiar-se* (from *mal* and *fiar-se* 'to rely on'), when they are followed by adjectives and verbs. On the other hand, there are nouns with a first component *mal*: *malnom* 'nickname' (from *mal* 'bad' and *nom* 'name') or a noun (*mal* 'evil'), where *mal* is not considered a prefix. Thus, *ben-* and *mal-* are sometimes considered prefixes, sometimes not.

En- is listed together with adverbs and prepositions whose distribution has been classed as prefixal, as well as with true prefixes. It is impossible, then, to tell whether it is entered as a true prefix or as a preposition in words like those in (2).

- (2) en-sucrar
 ?in sugar-INF 'to put sugar in'
- en-fosquir
 ?in dark-INF 'to darken'

However, for the following examples, Fabra claims that *en* is in fact the clitic *en*.

- (3) em-portar-se
 EN take-INF SE 'to take with'
- en-riure's
 EN laugh-INF SE 'to laugh at'

He justifies this analysis in view of the relation of these verbs with root verbs as shown in (4). In these examples *en* does not cooccur with a phrase introduced by *de* 'of' (phrases introduced by *de* cliticise in *en*):

- (4) Va riure's del seu pare moderadament.
 s/he-laughed-SE of her/his father moderately
 'S/he laughed at her/his father moderately'
- *Va enriure's del seu pare moderadament.
 s/he-EN-laughed SE of her/his father moderately
 (same meaning)
- Capses? Va emportar-se'n quatre.
 Boxes? s/he-EN-took SE EN four
 'Boxes? S/he took four'

He does not discuss the data in (4) nor make his analysis explicit; I shall come back to this issue in Chapter Six, for the moment I shall point out that I do not agree with all his grammaticality judgements.

The basis of Fabra's subdivision of the second class, "other compounds", is mainly syntactic: the first group (illustrated in (5)) corresponds to compounds whose first element is a verbal form, while in the second group (illustrated in (6)) the first element is nominal or adjectival. Coordinated compounds and "synthesised" compounds, illustrated later, also belong to the class of "other compounds".

- (5) [pical]_V [porta]_N
 knock door 'door-knocker'
- [alça]_V [prem]_V
 lift press 'lever'
- (6) a. [cama]_N [llarg]_A
 leg long 'long-legged'
- un ocell camallarg 'a long-legged bird'
- un camallarg 'a wading bird'

- b. [pit]_N [negre]_A
 chest black 'turnstone'
- c. [cap]_N [alçar]_V
 head raise-INF 'to raise the head'
- [car]_A [vendre]_V
 expensive sell-INF 'to sell at an expensive price'
- [llamp]_N [ferit]_{pp}
 lightning struck 'struck by lightning'
- [ver]_A [semblant]_{prp}
 true seeming 'likely'

Fabra classifies not only on the basis of the syntactic categories that can form a compound but also on some of their internal features, such as those derived from agreement: (6a) displays no internal agreement since *cama* is feminine and *llarg* is masculine, while (6b) supports agreement since *pit* and *negre* are both masculine. In (6c) Fabra considers verbal compounds with a verb in second position in its infinitive form, and also in participial (past and present) form. Such verbal and deverbal compounds (in his own terminology) are "composts en què el primer component és un complement del segon" ('compounds in which the first component is a complement of the second' Fabra 1956:152); the syntactic relation of *picaporta* in (5) with the phrase *picar portes* 'to knock doors' is also mentioned (Fabra 1956:151).

Compounds based on coordination of any two units of the same category, which are exemplified in (7), are also grouped as compounds without prefixes.

- (7) a. [figa]_N [flor]_N
 fig flower 'weak person'
- b. [sord]_A [mut]_A
 deaf mute
- c. a mata-degolla
 to kill throat-cut 'desperately'
- d. piu-piu (noise made by poultry)
 xiu-xiu 'whispering'
 ziga-zaga 'zig zag'

All examples in (7d) are onomatopoeic, and without reduplication would make no sense at all (**xiu*, etc). By including (7d) in the list of compounds, Fabra assigns the first member of the word (which is often identical to the second) to the set of words, too.

Also among "other compounds", there is a class which corresponds to compounds

"originats per la sintetització d'un conjunt de mots format segons les regles sintàctiques ordinàries" ('originating in the synthetisation of a group of words formed according to the normal syntactic rules' Fabra 1956:155). It is not clear whether, by synthetisation, Fabra refers to the impossibility of interpreting these compounds analytically (or, we could say, compositionally), or to their wordness (as opposed to phrasal character).

- (8)

aiguardent		
[aigua] _N [ardent] _A		
water burning		'liquor'
[mitja] _A [nit] _N		
mid night		
[tres] _A [peus] _N		
three feet		'trivet'
[cul] _N [de] _P [sac] _N		
bottom of bag		'cul-de-sac'

Their inflectional paradigms vary considerably. Observe e.g. the case of plural markers (the plural marker is -s):

- (9)

a. targeta postal		
card postal		'postcard'
targetes postals		'postcards'
b. ull de poll		
eye of chick		'callus'
ulls de poll		'calluses'
c. blau marí		
blue marine		'navy blue' [+PL] or [-PL]

Other nouns included here and formed following normal syntactic rules are:

- (10)

cap-i-cua		
head and tail		'reversible number'
ave-maria		
hail mary		
no-m'oblidis		
no me forget		'forget-me-not'

In a note (Fabra 1956:154) Fabra compares such nominal compounds to similar verbal "locutions" such as those in (11), without specifying how far the similarity goes.

- (11) fer denteta
make tooth-DIM 'to tantalise'
- anar de tort
go of askew 'to stagger along'

It is unclear whether Fabra wants to consider them compounds as well or not. We face here the problem of defining idioms.

Finally, the last group includes all the so-called "locucions adverbials, prepositives i conjuntives" ('adverbial, prepositional and conjunctive locutions' Fabra 1956:158): numerous adverbs, prepositions and conjunctions made of "conjunts més o menys sintetitzats" ('more or less synthesised groups' Fabra 1956:158) of words:

- (12) a-les-hores
in the hours 'then'
- si-nó
if not 'but'

The only criterion that allows the distinction between the examples in (8) and (12) seems to be the open character of the first in opposition to the closed one of the second (since, in a given grammar, prepositions and conjunctions at least do not belong to a productive class).

The criteria that Fabra plays with are quite numerous: syntactic regularity, category membership, morphological behaviour (agreement, etc.), origin (Latinate vs. non-Latinate) and, perhaps, productivity. His classification is cryptic, however, in that he does not use labels for different classes, nor does he make explicit why he establishes them. A lot is left to the intelligence of the reader.

2.2. Badia

Antoni M. Badia i Margarit's *Gramàtica catalana* [Catalan Grammar] devotes a section in the second volume to word formation; his main source is Fabra 1956, which he follows very closely, though often he also refers to A. Marvà's (1934) *Curs superior de gramàtica catalana* [Advanced Course of Catalan Grammar].

Badia divides all compounding into two classes (Badia 1962:380):

"composició por yuxtaposició" ('compounding by juxtaposition'), "la que resulta de unir dos o més paraules sin alterar las leyes gramaticales (o sea, que se observan la concordancia, régimen de preposiciones, complementos verbales, etc.)" ('that which results from putting together two or more words without altering the laws of grammar (i.e. obeying agreement, government of prepositions, verbal complements,

etc.)' ... [and] ... "composición elíptica" ('elliptic compounding'),
 "llamada también composición propiamente dicha" ('also called
 proper compounding')

These elliptic compounds are characterised in negative terms, as violation of syntactic rules of grammar. That syntactic rules can be taken as basic seems to be an underlying assumption: e.g. *palplantat* 'stick-planted' 'still as a statue' is elliptic with respect to its paraphrase *plantat com un pal* literally 'planted like a stick'. We will see below that this syntax-dependent approach is found in the work of several other grammarians.

The detailed classification which is on the basis of mother and daughter categories goes as follows.

(13) I. Compounds with N, A and num

[N A]_N
 [N N]_N
 [A N]_N
 [num N]_N
 [N A]_A
 [A A]_A
 [A N]_A
 [num (CONJ) num]_{num}

II. Compounds with N or A and V

[V N]_N
 [V V]_N
 [A V]_V
 [A Vpp]_{Vpp}
 [A Vprp]_{Vprp}
 [N V]_V
 [N Vpp]_{Vpp}
 [N Vprp]_{Vprp}

III. Compounds with categorially homogeneous members

[N N]_N
 [A A]_A
 [V V]_{loc}
 [x x]_{onom}

IV. Compounds with categorially different members

[N P N] _N	
P ∈ { <i>de</i> 'of', <i>en</i> 'in'}	
[N CONJ N] _N	
CONJ = <i>i</i> 'and'	
[phrase] _N	Eg: <i>pare nostre</i> 'paternoster'
[Adv N] _N	<i>sempre-viva</i> 'everlasting
[Adv A] _N	flower'
[S] _N	<i>Déu-vos-guard</i>
	'God be with you'

I and II provide a classification on the basis of mother and daughter categories. III groups compounds made of elements belonging to the same category (as one of Fabra's classes), but this class largely overlaps with I and II. Furthermore, III and IV, according to their names, "compounds with categorially homogeneous

members" and "compounds with categorially different members", constitute another classification, and should be complementary (constituents must be either the same or different), but they are not. So, despite the fact that it follows that of Fabra (1956) in many details and, also, that it is a lot more explicit, this is not an exhaustive and exclusive grouping of elements, as Fabra's classification was.

Note also that no compound with a preposition as first member appears in the list above. Badia takes Fabra's position in classifying as prefixal words whose first member is a preposition or an adverb. This is even less natural when he recognises (Badia 1962:379):

"Más discutible es la diferencia entre la composición y el uso de prefijos (casi todos ellos son preposiciones o adverbios con significado conocido, tanto si contribuyen a la formación de derivados como si no intervienen en ella) (...). Pero nosotros no lo hemos hecho así [incluirlos como compuestos], tanto porque los prefijos, en su calidad de tales, no siempre coinciden con las preposiciones y adverbios como por el evidente paralelismo que ofrecen, así, prefijos y sufijos." ('More arguable is the difference between compounding and the use of prefixes (most of which are prepositions or adverbs with well-known meaning, whether they contribute to the formation of derivatives or not) (...). But we have not done it this way [include them as compounds], both because prefixes, as such, do not always coincide with prepositions and adverbs and because of the obvious parallelism that prefixes and suffixes then offer.')

So, Badia seems to be caught in Fabra's contradiction, too. What is worse, in his IV class one can find structures, [Adv N]_N and [Adv A]_N, exemplified by *sempreviva* 'everlasting flower' and [P N]_N, *acompte* 'payment', whose first members are an adverb and a preposition, respectively. In II we can also find a verbal compound type formed by an adverb followed by a participle (past or present): *proppassat*, of *prop* 'near' and *passat* 'past', 'recent', *propvinent*, of *vinent* 'coming', 'near coming'. These same two examples appear under derivation (Badia 1962:369), in the list of derivatives of the prefix *prop*. In consequence, the lack of clarity in the theoretical treatment of adverbs and prepositions has as effect the double classification of several words.

Another variation of Badia with respect to Fabra is that the former includes and the latter does not include adjectives such as those in (14) in his inventory of compounds of the form [A A]_A, despite the ungrammaticality of **greco* and **teòrico* (which contrasts with the grammaticality of *grec* and *teòric*).

- (14) greco-llatí
Greco Latin
- teòrico-pràctic
theoretico practical

The -o is a Latinate marker, not present, for instance, in (15).

- (15) diccionari grec-llatí
dictionary Greek-Latin

One may want to distinguish (14) from (15) without, of course, having to refer to the origin of both, but simply by looking at the vocabulary of Catalan at a given time. Including *greco-llatí*, together with *grec-llatí*, in the list of compound adjectives makes the distinction difficult.

Finally, in relation to Badia (1962), it is worth noting some of his observations on the inflection of compounds. There are several patterns of pluralisation of nominal compounds, as mentioned earlier. Badia notices alternations in [N N] nominals, for instance in *pruna clàudia* 'Claudia plum' 'greengage', the plural of which can be *prunes clàudia* or *prunes clàudies* (to be compared with e.g. *ferrocarril*, literally 'iron-rail', 'railway', *ferrocarrils*, **ferroscarrils*). He notes that there is a dominant tendency to the pluralisation of both elements in this type of compound (see Badia 1962:382).

2.3. Moll

Another grammar that addresses morphology is Francesc de B. Moll's *Gramàtica catalana -- referida especialment a les Illes Balears* [Grammar of Catalan -- with Special Reference to the Balearic Islands]. This was published in 1975, after an earlier version of 1968, with a didactic purpose. Compounds are the subject of the last, very short lessons, and prescriptive comments are added, as in Fabra 1956 and Badia 1962.

Moll lists Catalan compounds -- corresponding to the data that I have already mentioned -- in ten groups plus an extensive class of compounds based on prefixation (as Fabra did; there is no variation in this respect). The different lists are construed on the basis of daughter categories.

Alallarg 'long-winged' and *aiguarent* 'liquor, brandy' are kept in different classes, though, because, despite the fact that they are formed by a noun followed by an adjective, the first is an adjective and the second is a noun; so this is the only case where Moll claims that, besides reference to the daughter categories, reference to the mother category is necessary.

Within the group of adjectives of the form [A A], he distinguishes the subgroup of adjectives exemplified in (14) above, whose first constituent ends in the Latinate borrowing -o (as in *hispano-romà* 'Hispano-roman', with *hispano* instead of *hispanic* or *espanyol* 'Spanish', or *russo-japonès* instead of *rus-japonès* 'Russian-Japanese'). Moll also defines a class for compounds such as the following, which result from

derivation on a compound².

- (16) a. lliure-canvista
 free change-ist 'free trader'
 &canvista
 chang-ist
 lliure canvi
 free change 'free trade'
- b. bon-homia
 good man-hood 'good nature'
 ?homia
 man-hood
 bon home
 good man 'good-natured man'

Moll points out an interesting variation between continental and Balearic Catalan which occurs in [N A] adjectives: in continental dialects the adjective agrees with the compound-external noun it modifies, while in Balearic Catalan the adjective agrees with the compound-internal noun. The following examples of Moll illustrate the pattern for gender.

- (17) La nena està cap-baixa. (Continental Catalan)
 the girl is head-MASC low-FEM
 'The girl is sad'
- El noi és un llengua-llarg.
 the boy is a tongue-FEM long-MASC
 'The boy is a chatterbox'
- (18) La nina està cap-baix. (Balearic Catalan)
 the girl is head-MASC low-MASC
 'The girl is sad'
- S'al·lot és un llengua-llarga.
 the boy is a tongue-FEM long-FEM
 'The boy is a chatterbox'

Note that there are nouns of the form [N A] which, both in Balearic and in continental Catalan, obey the pattern displayed here by the Balearic adjective:

- (19) un pell-roja
 a-MASC skin-FEM red-FEM
 'a red-skin, an American Indian'

Moll (1975) mentions for the first time the problem of determining for verbal compounds what form of the verb is present. According to him, *pica* 'knock' in

²The symbol "&" before a word means possible but non-existent; the notation follows Roeper and Siegel 1978. I shall use this diacritic only when I find it strictly necessary.

picaporta 'door-knocker' corresponds to the second person singular imperative (which is homophonous with the third person singular of the present indicative). Likewise, following Moll, *gira* 'turn' and *volta* 'rotate' in *giravolta* 'rotation' are imperatives, though he considers that the verbs which appear in [V CONJ V] compounds might be in indicative (i.e. third person singular of the present indicative). Unfortunately, the choice is not argued for in any way.

2.4. Ferrater

Gabriel Ferrater's "La composició nominal" [Nominal compounding] was published as an article in 1970 and posthumously as Ferrater 1981b; as the title suggests, this paper refers only to nominal compounds. The basic classification is taken from Benveniste (1974a, 1974b), but extended in the light of Bally 1950 and the work of Jespersen in general. The goal of the article is (in terms of Benveniste that Ferrater makes his own) "estudiar cada tipus de compostos com la transformació d'un tipus d'enunciat sintàctic lliure" ('to study each type of compound as the transformation of a type of free syntactic proposition' Ferrater 1981b:55). The goal is new, Ferrater says, in a grammar of Catalan, given that the previous grammars have not been developed in this spirit. Ferrater's classification is interesting not only from the point of view of syntax, but also semantically; in fact, it is probably syntactically perspicuous because it is semantically relevant. It is useful to give the classification in detail and to consider the semantic properties of each class:

- Dvandva type

These are conjunctions of two nouns or adjectives, with a meaning which is the addition of the two meanings:

Austria-Hongria
Austria Hungary

pèrdues i guanys
losses and gains

sord-mut
deaf mute

- Non-additive conjunction

These are exemplified by the following:

nord-est	
North East	'somewhere between North and East'

est-oest	
east west	'from East to West'

- Metaphoric modification

The first noun is the head of the construction, and the second modifies it, but the modification is metaphoric; that is, given a compound [X Y], its interpretation is 'X which is like a Y':

paper moneda
 paper coin
 'paper to be used like coins, bank note'

col-i-flor
 cabbage and flower
 'cauliflower'

- (Non-metaphoric) modification

These consist of a noun modified by an adjective:

estat major
 state major 'military staff'

curt circuit
 short circuit

A problem arises with this type in distinguishing it from free phrases; the criteria to distinguish it are those of Bally (fixed order, no possibility of modifying only one of the components).

- Synaptic group in Benveniste's terminology

tauleta de nit
 table-DIM of night 'bedside table'

bany-maria
 bath mary 'bain-marie'

carrer Muntaner
 street Muntaner

The criteria to keep these apart from phrases are again the ones of Bally mentioned above. A subclass of synaptic compounds omits the preposition; the cases without preposition derive historically from Latin phrases whose second element was marked as genitive.

- [V N] compounds

This class is exemplified by the following:

passa-port
 pass port 'passport'

where the noun corresponds in general to the direct object of the verb, with a few exceptions (as *caga-ferro*, literally 'shit-iron', 'clinker', where, as in a passive, the noun corresponds to the agent). Disagreeing with Moll, Ferrater considers the verb to correspond to a sort of basic form (after Jespersen) or to the root form (cf. Bally 1950); this idea is supported by Ferrater's example *cobre-llit* literally 'cover-bed', 'bedspread', where *cobre* does not coincide with any (other) realised form of the paradigm of *cobrir* 'to cover'.

- Bahuvrihi type

These are exemplified by the expressions:

pell-roja
 skin red 'red skin'

pit-roig
chest red 'red-chested'

cap de trons
head of thunder 'scatter brain'

with two levels of predication (while all the compounds above had one):
predication of quality and predication of attribution (e.g. 'the skin is
red' and 'the skin belongs to somebody').

- Coordinated bahuvrihi type

The compounds with two levels of predication but conjunctive shape
can be exemplified by:

plats-i-olles
dishes and casseroles
'shop where you buy dishes'

Ferrater leaves aside some nominal compounds, so that his classification is not exhaustive, though more or less representative of the syntactic possibilities of compounding in Catalan. Some specificity is missing; for instance, he pays no attention to syntactic characteristics such as the presence or absence of conjunctions. Ferrater himself writes that his inventory is far from original, but it includes novel details as far as the semantics of compounds is concerned. In fact, the question is whether we expect a grammar to give so much detail or rather leave some of the distinctions drawn above as linguistically underdetermined.

2.5. Wheeler

Next I shall look at the work of Max W. Wheeler on compounds, which concerns their phonology and their syntax; it is the first approach carried out unequivocally in a transformational framework. The phonological aspects are developed very much under the influence of *SPE* (Chomsky and Halle 1968). The syntactic analysis suggests deep structures for compounds in an early transformational style.

Wheeler takes "word" as a primitive of the grammar: words are elements dominated by the nodes of noun, adjective, verb, adverb, preposition or pronoun (with some exceptions irrelevant to the problem at issue). Wheeler 1979 and more briefly Wheeler 1977 present the rules necessary to predict the stress patterns of Catalan; I shall not enter the discussion on the rules of stress assignment, which would take me too far from the topic of this thesis, and shall just consider a subsequent phonological rule particularly relevant for compounding. Compound words, like phrases, make a stress reduction rule necessary, a rule which is applied cyclically.

- (20) Within [...] reduce all stresses except the rightmost
by one.
(Wheeler 1979:48)

([...] indicates a phonological domain). The rule expresses his observation that "compound words (..) have, with a few exceptions, a stressed vowel in each part, the rightmost stress of the whole being the primary stress" (Wheeler 1979:47) (consider e.g. *parallamps* [pàrəláms], where the first *a* bears a reduced stress). Exceptions are found such as compounds with one stress only and vowel reduction in the first member of the compound (e.g. *només*, literally 'no more', 'only' [numés]), or alternations where the application of these rules appears to be optional or varies across speakers (e.g. *entreobrir* 'to half-open'). It seems that a characterisation of compounding in phonological terms is difficult: no phonological feature is true of all compounds and only them.

In relation to derivation, Wheeler finds prefixes which keep their own stress when they enter the formation of a word; for example, *neo-* 'new' and *post-* 'after', both Latinate. To capture their behaviour in accordance with the stress rule in (20), the move proposed in Wheeler 1977 is to assign them a word boundary (like the elements that belong to a compound) and a syntactic category (noun or adjective, though we do not know on what basis). Their exceptionality would be, of course, that they have to be marked in order not to be inserted in a phrase structure as independent words (if they only become independent words by apocope: *auto* from *automòbil* 'automobile' and so on).

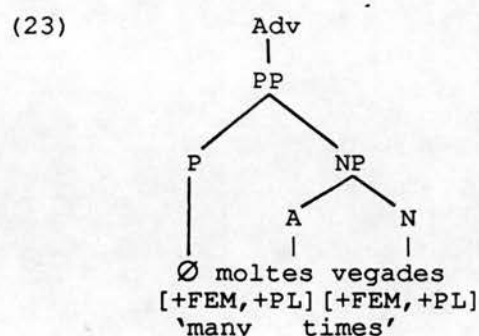
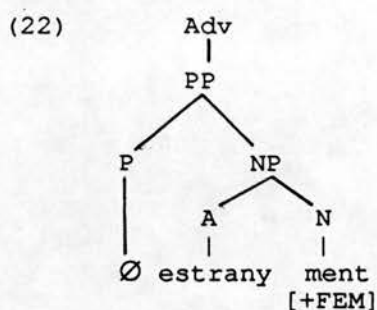
Other phonological rules are mentioned (Wheeler 1977:245) which apply quite regularly between the two constituents of a compound the same way as they apply between independent words -- they are rules of consonantal and vocalic sandhi. Curiously enough, one of the rules (devoicing of final obstruents) seems to apply after the prefix *sub* 'under' (another Latinate form, which does not occur as a free word: *sota*/**sub la cadira* 'under the chair') so, once again, the class defined with a phonological criterion does not conform to compounds.

Wheeler's treatment of the adverbs in *-ment* '-ly' as compounds is also motivated, in part, by their phonological peculiarity: the presence of two stresses, the rightmost one being primary (as in compounds in general). Another property of the adverbs in *-ment* is the appearance of the feminine form of the adjective at their base, instead of the unmarked form which would be usual in derivatives:

- (21) *estranya-ment*
 strange-FEM ly 'strangely'
- franca-ment*
 frank-FEM ly 'frankly, sincerely'

Both characteristics are exceptional if we analyse these adverbs as the result of

suffixing *-ment* to an adjective to obtain an adverb. So Wheeler postulates the analysis in (22), parallel to the analysis of a zero-preposition adverbial in (23) (Wheeler 1977:239).



That is, *estranyament* would be "un sintagma nominal: adjectiu + nom de gènere femení, potser dins d'un sintagma preposicional amb preposició zero" ('a noun phrase: adjective + noun of feminine gender, perhaps in a prepositional phrase with a zero preposition' Wheeler 1977:239). The author calls this a lexicalised construction and it is not clear to me why it is a compound; *moltes vegades*, analysed the same way, is certainly not a compound; so, unless we consider that its being lexicalised makes it a compound, the analysis proposed suggests that it is a phrase. Notice that to insert *-ment* under the N node it needs to be an independent word; now, there is the feminine word *ment* 'mind', but, for the obvious semantic reasons, Wheeler prefers to resort to another lexeme *ment* meaning something like 'way, manner', which would surface in this class of adverbs and nowhere else.

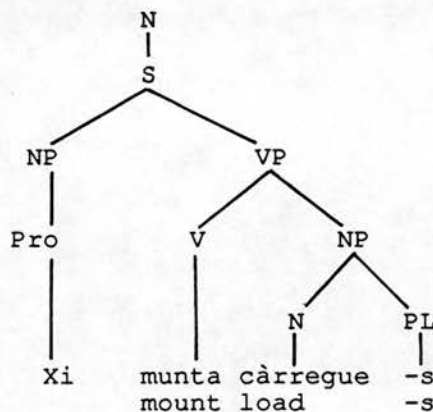
Finally, this approach is meant to allow the deletion of *-ment* in adverbs coordinated in pairs. The normal pattern of noun deletion (if the adequate analysis consists of a deletion) operates on the first and not the second noun, as shown in (24a). However, *-ment* adverbs can be deleted in either the left (popularly) or right conjuncts (as prescribed), as shown in (24c) and (24d) respectively.

- (24) a. poques o moltes vegades
 a few or many times
- b. ?poques vegades o moltes
 a few times or many
- c. estranya i insospitadament
 strange and unsuspectedly
- d. estranyament i insospitada
 strangely and unsuspected

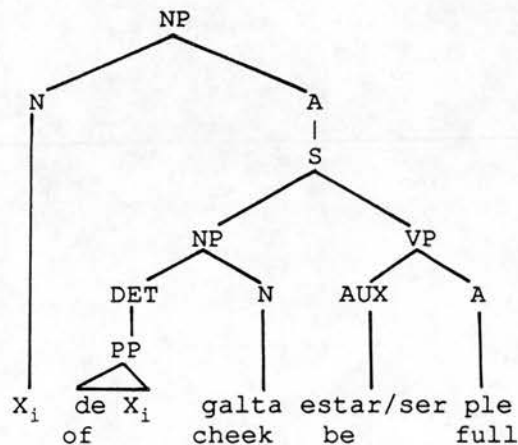
Wheeler argues that his analysis supports the grammaticality of (24d) through the normal processes of noun deletion; we note here that it supports at least as well (24c) (which he does not mention), analogous to (24a).

The other three kinds of compound analysed are exemplified in (25), where (25a) is a nominal compound, (25b) an adjectival one, and (25c) a verbal one. In all of them the element that surfaces on the right has the same category as the compound (the adverbs in *-ment* above are an exception to this, though not the only one: see *terraplè* 'soil-full' 'slope'). The deep structures in (25) are based on independently motivated phrase structures; what differentiates them from phrases is the presence of an X^0 projection mother node, of category N, A, or V. They are generated via $N \rightarrow S$, $A \rightarrow S$, and $V \rightarrow VP$, respectively.

- (25) a. munta-càrregues
mount loads 'lift'

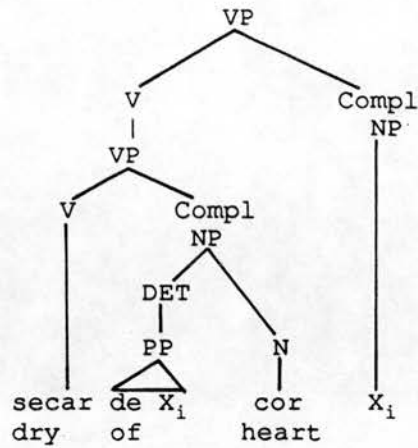


- b. galta-plè
cheek full



paraphrase: (X_i) de qui la galta està/és plena
of whom the cheek is full

c. cor-secar
heart dry-INF 'to wither'



paraphrase: secar (el) cor de X_i (a) X_i
dry the heart of X_i to X_i

These deep structures are the ones which are interpreted. The rules to derive compounds from these deep structures are not specified in this paper, but they would have to include deletions; as a consequence, unless conditions on deletion are established, recoverability is lost (this is not surprising given the date of publication of Wheeler's work). Second, in our present perspective, the analyses proposed violate the X-bar principles; though this was not considered a problem at the time, we might want to avoid it in our analysis.

2.6. Grossmann

Maria Grossmann published a paper on the *mntacàrregues* type of compound (see (25a) above) in 1986; it is the only paper which takes E. Coseriu's approach. It classifies a corpus of 600 compounds she has found in several Catalan dictionaries; thus she has classified words as they are in the actual vocabulary of Catalan (the norm in Coseriu's (1981) terms), not as they could be according to the grammar of Catalan only (i.e. according to what Coseriu calls the system of the language). Coseriu's notion of compound does not coincide with anybody else's notion of compound that I know of, and is much wider than the one commonly held, which does not include e.g. *dreamer* among compounds. Compounding is characterised as the union of two elements of the base, linked by a grammatical determination (Grossmann 1986:155, from Coseriu 1978). When one of the units is not identifiable with a free word of the language, the composition is called prolexematic; this type of compounding is exemplified by *somniador* 'dream-er' 'one who dreams'. When the two units are free words, this kind of composition is called

lexematic, and can be exemplified by *filferro*, 'thread-iron', 'wire'³.

Having the overall approach in mind, let us turn to the analysis of *somiatruites* 'dream-trouts' 'visionary'. This compound results from, first, "prolexematic compounding" (giving *somiador* 'dreamer') and, second, the "lexematic compounding" of *somiador* and *truites*, through some mechanism involving deletion of the suffix in *somiador*, which does not surface in *somiatruites*. Compounds are attributed the meaning of underlying sentences, e.g. *somiador* is related to 'somebody dreams'.

The two elements that eventually surface in a compound like *somiatruites* are called determiner and determined. Compounds are found with the same elements placed in the inverse order, determined plus determiner, such as *camalluent* 'leg bright' 'with shiny legs'. The only consequence of this variation in order is that, while *somiatruites* is a noun (which can be adjectivised), *camalluent* is an adjective. (The reasoning behind this conclusion seems to be that *truites* determines, in some way, the category of *somiatruites*. I shall argue later that this is not so).

The compound can be the result of the nominalisation of the underlying predicate or subject -- though normally it is the second. The paraphrase that corresponds to a compound with a determined (subject) agent ((26a) below) is 'X+V+direct object or adverb', e.g. *parallamps* 'X that stops the lightning', 'lightning conductor'. When the determined element is the nominalisation of the predicate ((26b) below) the paraphrase is 'the fact of+V+direct object' in most cases (Grossmann 1986:164). (26b) is, however, very rare, and sometimes archaic. Very briefly, I summarise her classification with the "classes" she uses.

³Strangely enough, Grossmann forgets about the framework she has chosen quite soon, to refer to Coseriu's first kind of compounds as derivatives and to the second as compounds.

- (26) a. Agent
 [+animate, +human]
 names of jobs
 occupations
 characteristic designations
 e.g. *somia-truites* 'dreamer'
 [+animate, -human] [+animal]
 names of birds
 fish
 insects
 [-animal]
 names of mushrooms
 plants
 fruits
 [-animate]
 names of tools
 weapons
 parts of ships
 places where activities take place
- b. Predicate
 e.g. *besa-mà*
 kiss hand 'fact of kissing the hand'

In e.g. *guarda-roba* 'cloak-room' Grossmann designates the place where the action occurs as the "agent"; one has to suppose, then, that when saying "agent" she gives a very broad sense to this term.

The essential problem with Grossmann's classification, for the present purposes, is that it contains almost no linguistic information at all. Her classemes have not been proved to have any relevance in the organisation of the lexicon; at most, they are hyperonyms of the word they are meant to characterise. When she establishes the feature [+/-animal], one does not see any reason to stop there and not to have one for [+/-fish]; the fact that psycho-philosophical arguments have been adduced in componential semantics for a universal feature [+/-animal], but not for one [+/-fish] is not something that Grossmann records. So her classification of actual compounds of Catalan is statistically informative (in proportion, there are many compounds designating birds and insects, etc.) and this may have historical, or lexicographic interest, but not linguistic relevance.

Grossmann's interest is in the norm (that is, in the existing, with all its idiosyncracies) and not in the system of Catalan. This is the origin of her last question: Why is the norm as it is, given the relative freedom allowed by the system? This issue is independent, I think, of our present concern, namely the determination of what is possible, regardless of what has, accidentally or not, been.

2.7. Cabré and Rigau

Lexicologia i semàntica [Lexicology and Semantics], by Teresa Cabré and Gemma Rigau, was written a few years before its publication in 1986; it was written within transformational generative grammar but without reference to the Government and Binding framework. It is regarded by the authors as complementary to Mascaró 1986, published in the same series, despite differences in their treatments of some problems; I shall return to some issues in argumental structure of verbal compounds in section 2.8., when reviewing Mascaró 1986.

The work of Cabré and Rigau is on the syntax and semantics of compounds, so that their definition of compound is deliberately and explicitly non-phonological. The definition (Cabré and Rigau 1986:134) goes as follows:

"La composició (...) permet l'obtenció de peces lèxiques a partir d'elements que ja figuren al diccionari o lexicó. (...) Per composició aconseguim nous mots per l'adjunció de radicals. En català, els constituents d'un mot compost han de pertànyer a una de les següents categories gramaticals: Nom, Verb, Adjectiu i Adverbi." ('Compounding enables the acquisition of lexical items from other elements which are already in the dictionary or lexicon. (...) Through compounding we obtain new words by adjunction of roots. In Catalan, the constituents of a compound must belong to one of the following categories: Noun, Verb, Adjective and Adverb'.

After this definition, the possible combinations of the four categories are examined syntactically and semantically. I shall only present the innovative and/or contentious points of their analysis.

In the first place, one might ask what singles out the four categories purported to make up compounds. Adverbs have been included but prepositions are still treated along with prefixes. They argue that *sobre* 'on', for instance, functions in word formation the same way as *super* (the Latinate prefix with the same meaning), though *super* does not appear freely in sentences (*No vull els papers sobre/*super la cadira* 'I don't want the papers on the chair'); the parallelism of *super* and *sobre* is going to be expressed in the lexicon (since they postulate entries for affixes) by their coincident lexical information [+prefix, --R_V], where R_V stands for verbal root. The preposition *sobre*, which shares meaning and phonological shape with the supposed prefix *sobre*, has an independent entry. A factor that they do not consider, and which does not support their approach, is that another element of the same kind, *entre* 'between', does not have the same combinatorial possibilities of its Latinate counterpart *inter*, so that speakers cannot really consider such pairs as syntactically equivalent (and from the point of view of register they are recognised to be different).

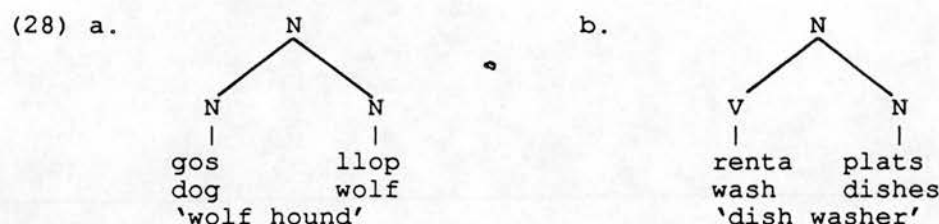
The adverbs in *-ment* are obtained by derivation⁴, so that *-ment* is a suffix and the fact that it attaches to feminine adjectives is simply specified in the subcategorisation frame of its entry.

The "proper compounds" of Catalan, on the other hand, are said to have the following structures (Cabr  and Rigau 1986:135)⁵:

- (27) [N N]_N
 [V N]_N
 [A A]_A
 [N A]_A
 [N V]_V
 [A V]_V
 [Adv V]_V

Some vacillation in category assignment (*primfilar* 'fine-spin' 'to split hairs' [A V] and [Adv V] in Cabr  and Rigau 1986:148 and 149 respectively) suggests that perhaps the distinction adjective/adverb is not clear enough.

Now, most compounds above are right-headed; headness is determined on syntactic grounds (the head's syntactic category is shared with the mother node), and semantic ones (the composite meaning is determined primarily by the head); also, morphological markers tend to appear on the head. However, there are some cases of left-headness (28a) and of exocentricity (28b):



In (28a) a sign of left-headness is the plural form *gossos llop*, **gos llops*; but there are counterexamples to this, such as *filferro* 'wire', whose head is on the left but marks its plural on the right -- semantically a *filferro* is a kind of *fil* 'thread', not of *ferro* 'iron'. In (28b), the compound as a whole does not inherit the syntactic and semantic properties of the right-hand noun, but of the subject of the verb, etc. Cabr  and Rigau observe that the verb must correspond to a verb of action, so that *semblar* 'to seem', *estar* 'to be', and all stative verbs are excluded from the first position.

Some other interesting remarks on semantics are made about the relation holding

⁴Cf. Espinal 1990a, 1990b.

⁵A *pit-roig* 'chest-red' 'robin' type, of [N A]_N structure is derived by a process of semantic transfer (Cabr  and Rigau 1986:144, taking the concept from Leech 1974).

between the two elements of a compound; when these elements are nominal or adjectival, no synonyms enter the construction of a compound (**lusitano-portuguès* 'Lusitano-Portuguese'), nor complementary words (**vida-mort* 'life-death'), and hyperonyms and hyponyms only rarely (so **catalano-gironí* 'Catalan-from Girona', because 'from Girona' implies 'Catalan'). Antonyms, however, can be combined, e.g. *compra-venda* 'purchase-sale' 'exchange'.

On syntactic grounds, a group of compounds is excluded from (27): the ones called synapctic in Benveniste 1974b. This group includes structures already generated by a phrase structure grammar of Catalan, like [N [P N]_{PP}]_N (*pa de pessic* 'bread-of-pinch' 'spongecake'), [N A]_N (*estrella polar* 'star-polar' 'Pole star'), [N CONJ N]_N (*col-i-flor* 'cabbage-and-flower' 'cauliflower'), plus adverbial phrases (*tothora* 'all-hour' 'always'), and prepositional phrases (*a la vora de* 'to the edge of' 'near'). Together with these, there are compounds in which the mother category is not the category of any of the constituents (e.g. *puja-i-baixa* 'go up-and-go down' 'action of going up and down') and "apparent verbal phrases which behave as a lexical element" (Cabr   and Rigau 1986:149), like *somiar truites* 'dream-trouts' 'to fantasise'. This class includes compounds with normal phrasal order or presence of a preposition or a conjunction, plus words which can be generated by syntactic rules if category changing rules are used. However, it is not clear where to draw the line between this class and proper compounds on the one hand, and idioms on the other. Why is *somiar truites* a synapctic compound while *anar de tort* 'to stagger along' is an idiom (Cabr   and Rigau 1986:151)? Or why are some words whose constituents appear in the normal phrase order proper compounds while others are not? The criteria to distinguish synopsis from phrases are:

1. Morphological: (i) Impossibility of making a pause in the middle of the compound, so that no other word can be inserted in that position (*pa de pessic* but **pa petit de pessic* 'small bread of pinch'); (ii) Formation of the plural which does not have to coincide with the normal pattern of pluralisation (*col-i-flors* but **cols-i-flors* 'cauliflowers').
2. Semantic: Synapctic compounds denote a unit of image or idea (see Bally 1950). This criterion is not sufficient, as the authors point out (furthermore, it is not clear what unit of image or idea means).
3. Syntactic: (i) No coordination of one element of the compound is possible (*esperit de vi* 'spirit of wine' 'alcohol' but **esperit de vi i de 90 graus* 'spirit of wine and of 90 degrees'); (ii) Modifiers can only apply to the compound as a whole (**esperit de [vi negre]*, literally 'spirit of red wine'); (iii) Compounds are anaphoric islands; as a consequence, it is impossible to substitute only one of the terms of the compound by an anaphoric pronoun:

- (29) *Aquí hi ha un molí d'aigua que brolla abundantament.
Here there is a mill of water which flows abundantly
'There's a watermill which flows abundantly'

Most of these criteria are relevant, but as Cabré and Rigau say, not sufficient to establish a class of compounds separate from the class of phrases, so one may want to conclude with Bally (1950) that perhaps such a class does not exist in the language itself.

On the whole, this piece of work has the advantage of not falling into syntactic reductionism; the authors consider word formation a truly independent way of conveying meaning. Only the educational purpose of the work prevents it from being more formal and, therefore, its explanations as they stand have little predictive power.

2.8. Mascaró

Before his *Morfologia* [Morphology], published in 1986, Joan Mascaró had already studied compounds in an unpublished paper on non-main stress in Catalan (Mascaró 1975); an experiment on stress in compounds is reported in Mascaró 1983, and it is this work to which I have had access. Contrary to the hypothesis (first expressed in Fabra 1912) that compounds have secondary stress, together with main stress, Mascaró argues that his experiments show the non-existence of a secondary stress in many compounds. His subjects were presented pairs of homonymous utterances varying only in that some words exhibited secondary stress while others did not; a third set of utterances included compounds which the subjects were required to classify as of secondary stress or no secondary stress. The result of the experiment was that compounds were grouped together with words without secondary stress. Therefore, perceptually, (at least some) compounds do not have a secondary stress before the main stress. Mascaró's assumption for this argument to hold must be that perception mirrors the facts of production in all respects. Accordingly, in Mascaró 1983 the majority of compounds are subject to a rule of deaccentuation. I shall not discuss in any more detail the proposals of Mascaró on the phonology of compounds, which would demand the development of his phonological theory. Mascaró's view differs in two main respects with Wheeler's: (i) Stress is not assigned by rule, while it is in Wheeler's model; (ii) Compounds do not present secondary stress, while they do for Wheeler.

Mascaró's *Morphology* is a complement to Cabré and Rigau 1986 in giving phonological depth to the study of Catalan compounds. For a start, Mascaró's

definition of compound is basically and, in problematic cases, crucially phonological. The difficult distinction between derivatives, compounds and lexicalised phrases is resolved, finally, by adopting a phonological definition.

Let us start with the first distinction: derivatives versus compounds. The criteria available are phonological and morphological, since no semantic criterion is considered sound enough (Mascaró 1986:46). Phonologically, compounds present suppression of one subjacent stress without vowel reduction, while derivatives present vowel reduction. A third group of words with more than one realised subjacent stress (*paper moneda* 'paper-coin' 'bank note', *cap de turc* 'head of Turk' 'scapegoat') are not considered as compounds but lexicalised phrases. Other phonological phenomena which occur in phrases also take place between the constituents of compounds (as pointed out by Wheeler, see section 2.5.), but not inside a derivative. They are: (i) Deletion of final *n*: *mà* 'hand', *màllarg* 'long-handed', but *maneta* 'hand-DIM'; (ii) Deletion of final *r*: *flor* [flɔ], *flor de neu* [flɔðənew] 'flower of snow' 'edelweiss', but *floreta* [flureta] 'flower-DIM'; (iii) Simplification of clusters of final consonants: *front* [frɒn] 'forehead', *frontample* [frɒnɒmplɔ] 'broad-foreheaded', but *frontet* [fruntet] 'forehead-DIM'; (iv) Devoicing of final obstruents: *sud* [sut] 'South', *sud-est* [sutest] 'South-East', but *sudista* [sudistə] 'Southern'; (v) Voicing of final fricatives followed by a vowel (*braç* [z+V] 'arm', *braçample* [z+#] 'broad-armed', but *bracet* [brəset] 'arm-DIM' (Mascaró 1986:45). All this suffices to reach a phonological characterisation of compound (Mascaró 1986:46).

Morphologically, derivatives have their heads on the right (though this is not exceptionless: see the comments in section 4.4.2.), while compounds can have them on the left or have splitting heads. Morphologically also, compounds present inflectional markers on the right while lexicalised phrases present them following the usual syntactic rules of inflection (*cap de turc*, plural: *caps de turc*, **cap de turcs*). Binary branching is also regarded as characteristic of (though not exclusive to) compounds (Mascaró 1986:76).

These criteria determine the classification of certain words as compounds, once the Principle of Indirect Justification, which Mascaró (1986:19) maintains, is accepted. This principle asserts that

"X és un radical si forma part d'un mot YXZ i Y, Z són altres radicals o afixos, encara que X no sigui d'alguna manera o altra un "mot independent". Així, tot i que no hi ha un mot *mor(a)* que s'hi relacioni, *mor* és una arrel que apareix a *demora*, *demor*, *demorar* (...). Acceptem aquí, per tant, el principi de la justificació indirecta, aplicat a qualsevol categoria" ('X is a root if it is part of a word YXZ and Y, Z are other roots or affixes, even if X is not, in one way or another, an "independent word". So, even if there is no word *mor(a)* 'pone' related to it, *mor* 'pone' is a root which appears in *demora* 'postpone',...(...). We accept

here, then, the principle of indirect justification, applied to any category'.

As a consequence, for instance, Mascaró takes the adverbs in *-ment* to be compounds, contra the treatment of Cabré and Rigau (1986).

Compounds are characterised as having two roots (while derivatives have one) (Mascaró 1986:22) or as the result of relating one lexical element with two or more lexical elements (at the base); leaving aside the vagueness of the terminology, a problem arises because of the adoption of the Principle of Indirect Justification. The only properties that unambiguously distinguish derivatives and compounds, for Mascaró, are phonological; they are the only empirical basis of the distinction. However, there are prefixes with a vowel which, like the vowel of the first element of a compound, is not subject to vowel reduction; given the Principle of Indirect Justification I see no way of telling the resulting derivatives apart from compounds. So, if we still want to keep these words in the group of derivatives, we have to find a more complete definition of compound or, alternatively, we have to consider the Principle of Indirect Justification untenable.

Mascaró states that his criteria, when applied to the existing vocabulary of Catalan, do not cluster together, so that it is impossible, for instance, to come to a "binary classification of complex words into derivatives and compounds" (Mascaró 1986:44). Examining the category of the constituents of an expression that might be a compound, the morphological behaviour of the expression and its phonological properties gives diverging classifications for the expression.

Mascaró claims that there is a phonological restriction applying to the $[V N]_N$ compound type. The first element, the verb, which corresponds, according to him, to the third person singular of the present indicative -- the least marked form of the conjugation -- can only end in a non-stressed vowel, and not in a stressed vowel, a consonant or a semivowel. In effect, one can find words with a vowel that does not appear in the third person singular present indicative, to avoid the first word ending in a consonant: *baticor* 'beat heart' 'palpitations', *beat* third SG pr IND being *bat* instead of *bati*. However, there are a few cases with a final consonant in the first word (Mascaró 1986:61) and for the production of compounds of this kind I do not agree with Mascaró on his judgements: **treu-taps* 'take off-corks' 'corkscrew', **protegeix-dits* 'protect-fingers' 'finger-guard' seem right to me. So, even if most compounds of this type have as their first constituent a verb of the first conjugation (with third person SG ending in a non-stressed vowel), the argument is not conclusive and at most we can talk of a tendency in the existing vocabulary.

Let us turn now to the syntactic features of compounds. Mascaró's classification is made on the basis of headedness:

1. *Parallamps* 'lightning-conductor' type: no head
2. *Barba-roig* 'red-bearded' and *terratrèmer* 'to quake' type: right headed
3. Syntagmatic compounds: head (right or left) as in a phrase; this includes symmetric compounds (dvandva type), with splitting heads.

The most revealing part of Mascaró's analysis relates some of the types above with syntactic structures. The *para-llamps* 'stop-lightning' 'lightning conductor' type is a compound with interpretation $\{x \mid x \in A_1 \text{ and } P(A_1, A_2)\}$ where *parar* 'to stop' is a verb P that takes two arguments A_1 and A_2 , and *llamps* corresponds to A_2 , the object of *parar* (Mascaró 1986:61).

- (30) *escalfa-cadires*
 heat-up seats
 'somebody who outstays his/her welcome' (AG)
- escalfa-llits*
 heat up beds
 'bed-warmer' (INSTR)

That A_2 does not necessarily correspond to the subject of the verb is proved by the possible paraphrases of the compounds at issue:

- (31) *trenca-colls*
 break necks
 'death trap'
- *Aquest lloc et trencarà el coll.
 this place you it-will-break the neck
 'This place will break your neck'
- En aquest lloc t'hi trencaràs el coll.
 in this place REFL in-it you-will-break the neck
 'You'll break your neck here'

That A_1 does constantly refer to the direct object of the verb is supposed to be proved by the impossibility of finding indirect objects in the second position of this type of compounds:

- (32) **agrada-nens*
 please children
- agradar [als nens]*
 please-INF to-the children
 'to please children'
- **parla-sords*
 talk deafs
- parlar [a sords]*
 talk-INF to deaf-PL
 'to talk to the deaf'

There are some compounds, though, where it is, for different reasons, difficult to identify the noun with a direct object; *ferro* is the agent in (33a), *sol* in (33b) cannot correspond to the direct object without the compound being pragmatically peculiar, the compound in (33c) is hard to interpret because it is a gallicism, and that in (33d) because it derives from an imperative clause.

- (33) a. caga-ferro
 shit iron 'clinker'
- b. gira-sol
 turn sun 'sun flower'
- c. torna-vís
 return screw 'screwdriver'
 (from French *tourner* 'to turn')
- d. salta-martí
 jump martin 'tumbler'

Other compounds which have to be similarly interpreted as exocentric have the form [P N]_N, e.g. *entre-acte* 'between-act' 'interval, period of time between acts'.

Cara-girar 'face-turn-INF' 'to turn the face' (metaphorically 'to change opinion') includes the noun *cara*, an inalienable part of the body (the *barba-roig* 'red-bearded' type also includes an inalienable part of the body, as already remarked in Wheeler 1977 and Cabré and Rigau 1986). Mascaró claims that this compound type also involves the inclusion of the direct object of the verb *girar*. This triggers promotion of other complements of the verb, e.g.

- (34) a. Trenca les ales dels ocells.
 breaks the wings of-the birds
- b. Alatrencia els ocells.
 wing-breaks the birds
 'S/he breaks the birds' wings'

However, the proper paraphrase of (34b) is not (34a), but rather *Els trenca les ales als ocells*, with the preposition *a* 'to', not *de* 'of'. (34a) does not imply the inalienable possession reading of (34b).

Mascaró analyses many other compounds in terms of "incorporation" (Mascaró uses the term in a loose, pretheoretical sense) of complements to the verbs, as in (35).

- (35) coll-portar
 neck carry-INF
 'to carry on one's shoulders'
- portar a coll
 carry-INF on neck
 (same meaning)

(He also perceives deverbal compounds like *esmaperdut* 'mood-lost' 'disconcerted' as derivatives of a verbal compound & *esmaperdre* 'to disconcert', contra Fabra and most other linguists).

Finally, his description of verbal and nominal compounds with *entre* 'between' and *contra* 'against' sometimes rests on the notions of reciprocity and reflexivity respectively:

- (36) *entre-mirar-se*
 between look-INF-SE 'to look at each other'
- contra-atacar*
 against attack-INF 'to counterattack'
- contra-projecte*
 counter project 'project against another project'

However, not all compounds of this form are interpreted as reciprocals, e.g. *entreveure* 'between-see', 'to see faintly'.

Although Mascaró uses terms such as "argument" and occasionally mentions theta roles, his approach takes grammatical functions as primitives. This is not so in Cabré and Rigau 1986, which I turn back to now. According to them, in the *rentaplats* 'wash-dishes' 'dish washer' type, the noun in [V N] fulfils the semantic argumental requirements of the verb. *Rentar* 'to wash' requires a THEME, realised in *plats* 'dishes', as well as an AGENT or an INSTRUMENTAL and, in fact, the compound as a whole designates an INSTRUMENTAL, but could equally well designate an AGENT (the grammar does not tell us about the existing, but about the potential). Again following Cabré and Rigau 1986:142, some verbs like *guardar* 'to keep' have alternative argument structures -- in this instance [AG, TH, (LOC)] or [LOC, TH]; consequently, *guarda-roba* 'keep clothes' can correspond to the AGENT in the first theta-grid ('somebody who keeps clothes') or to the LOCATIVE in the second theta-grid ('cloak-room'). Not all nouns in such compounds are THEMES; e.g. this does not seem the case in (33) above. Other compounds, such as *passatemps* 'pass-time' 'entertainment' and *trencaclosques* 'break-heads' 'puzzle', involve a causative verb, so that *passatemps* can be paraphrased as 'alguna cosa que fa passar el temps' 'something that makes time pass', in which case the noun remains the THEME of the verb.

The only other compound type analysed from the point of view of argument structure is the one exemplified by *caragirar* 'to change opinion', where the noun *cara* 'face' corresponds to the THEME of *girar* 'to turn' (but not necessarily to the object of the verb; this THEME can be realised as the subject, too). Nevertheless, Cabré and Rigau associate the noun with a THEME, but do not exclude association with other theta roles. They also point out that the features of the verb are

transmitted to the compound node, so that, for example, *glaçar* 'to freeze', when transitive, is a causative verb and so will be *sangglaçar* 'to blood freeze'. The description of another problematic example is as follows; *calcigar* 'to step on' has this argument structure: [AG, TH/PATIENT]; *peucalcigar* 'to foot step on' is a compound based on it, which includes the THEME, *peu* 'foot', though *peucalcigar* has the same argumental structure as *calcigar*: [AG, TH/PATIENT].

- (37) a. El cavall va calcigar el peu d'en Pere.
 the horse stepped-on the foot of DET Pere
 'The horse stepped on Pere's foot'
- b. El cavall va peucalcigar en Pere.
 the horse foot-stepped-on DET Pere
 (same meaning)

Briefly, it seems that only one THEME theta role is assigned in (37a), while two of them are assigned in (37b), though it is not specified what element assigns those two THEME theta roles. Thus the relation between these examples and theta theory is left unaccounted for.

2.9. Duarte and Alsina

Duarte and Alsina 1986 takes a historical perspective, aiming at reconstructing the development of Catalan from Latin, within a generative model. The authors' goal is, then, different from mine.

For the construction of a synchronic grammar, however, it is worth recording two criteria which they use to distinguish compounds from syntactic phrases: (i) Agreement does not have to hold between determinant and determined element (as has already been exemplified); (ii) No deletion of elements of the compound is allowed, even when they are iterated. (By the way, if this is so, one would have another argument to exclude the adverbs in *-ment* from compounding, since they allow deletion). A possible counterexample to (ii) which comes to mind is German compounding, which seems to allow deletion.

In relation to diachronic grammar, Duarte and Alsina notice that Latin did not usually resort to compounding as a means of word formation. Yet, some Latin structures have persisted as compounds in Romance such as, for instance, [Adv V]_V or [V N]_N. This last type (the *parallamps* type) derives, etymologically, from the imperative form of the verb followed by the noun (I have already presented alternative synchronic construal for this verbal form). Latin also had adjectives of the structure [A A]; when the two adjectives are identical, the resulting adjective is superlative or intensive, in Sardinian, Italian and the dialect of L'Alguer (cf. Blasco 1984:302-303 and Duarte and Alsina 1986:154) -- the authors do not tell us

whether this happened in Latin, and, surprisingly, do not mention that this pattern is very common in all dialects of Catalan, not only the one of L'Alguer: *petit petit* 'small small' means 'very small' and so on.

Duarte and Alsina list all kinds of existing compounds according to mother and daughter categories; these categories include relative pronouns and reflexive pronouns; e.g. *qual-se-vol* 'which-SE-want' 'any', that is a sequence of relative pronoun, reflexive pronoun and verb, forming an adjective. In some cases, this analytic procedure is certainly taken too far: e.g. the components of *qualsevol* can be identified by the speaker (though the pattern is not productive), but *plus-quam-perfet* 'pluperfect' is a borrowing, and not a compound of Catalan, since there is no such word as *quam* in Catalan.

The relevant part of their book on historical grammar, although containing a rich list of existing complex words, does not include any hypothesis as to structures available for compounding in Catalan or in Latin (this may not even have been their purpose) and is less committed, theoretically, than the rest of their work, so less open to discussion.

Of the issues raised by the various authors above, most will be pursued in the remainder of this thesis, especially comprehensively in Chapter Six. No original contribution will be made, though, with respect to two topics, namely lexical decomposition (cf. Grossmann) and the phonological properties of compounds (cf. Wheeler and Mascaró). To recapitulate, the following subjects will be considered in the framework chosen here: What compound types are found in Catalan? What are their characteristics, with regard to (i) syntactic category, (ii) semantic correlates, (iii) other syntactic features, such as argumental requirements, and (iv) morphological features, such as gender and number? Finally, how can the line be drawn between compounding and other phenomena such as lexicalisation?

Chapter Three

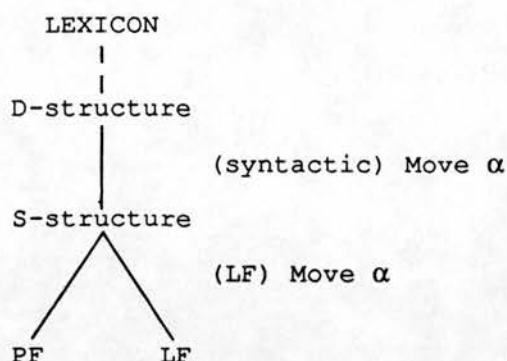
Compounding in GB; theories of meaning

3.1. Framework

Most probably because of the lack of a defined status of morphology in transformational generative grammar, the position of word formation has remained (with a few exceptions, cf. Borer 1984) unspecified in the later model, a model based on modules of principles (cf. Chomsky 1981, Chomsky 1986a). Most efforts to give an overall picture of word formation have consisted mainly in organising the lexicon (cf. Halle 1973, Aronoff 1976, and many others). It is no longer clear, however, that word formation and the organisation of the lexicon correspond exactly to the same area of the grammar. The general purpose of this chapter is, then, to see in what way we can account for various phenomena usually described as word formation, in the Government and Binding framework.

The elements that build up this kind of grammar are, basically, lexical entries on the one hand and, on the other hand, a set of principles which must be respected by the representations which the free application of a single rule, Move α , generates (together with simple concatenation). The kind and number of levels of representation postulated is justified by the empirical phenomena that the linguist wants to give an account of. I shall not reproduce the arguments used, but simply give a schema of the contemporary model.

(1)



It is desirable to keep the quantity of additional information minimal in the extension of the grammar to word formation. If the inclusion of new kinds of information does not prove necessary, the primitives on which our linguistic generalisations will be made are, firstly, the information already contained in the lexical entries, and, secondly, the principles of the GB subtheories: government theory, theta theory, Case theory, binding theory, bounding theory, control theory and X-bar syntax theory, together with Move α , if applicable. I shall look at these two issues in turn.

3.1.1. *Lexical entries*

Following Stowell 1981, the base rule component of the grammar and also the information on strict subcategorisation included in lexical entries are suppressed in the model employed here; the aim of this scheme is the avoidance of redundancy in the grammar. Both phrase structure rules and subcategorisation express the number and category of the arguments of a nucleus. The redundancy between the two has been often remarked, not only in the context of transformational generative grammar (cf. Gazdar et al. 1985). The absolute necessity of a subject (once phrase structure rules are eliminated) is indicated by the Extended Projection Principle of Chomsky 1981. Theta theory and subcategorisation are largely redundant with respect to one another: both include the number of arguments of a nucleus, but the theta-grid may include the subject, if it is assigned a theta role (in particular as an external theta role). The other piece of information that subcategorisation includes, the category of complements, is determined by Case theory together with Canonical Structure Realisation statements (cf. Chomsky 1986a) indicating the default syntactic categories realising theta roles. Thus, the function of subcategorisation and phrase structure rules is subsumed by elements of theta theory, Case theory, and Canonical Structure Realisation statements. The details are as follows:

- The number and semantic role of the complements of a verbal head

are determined by its theta-grid. I follow Williams 1981a in including the distinction between external argument and internal argument in a theta-grid. The first (in italics in the theta-grid) corresponds to a theta role assigned by the maximal projection of the verb, that is, by the VP, and, therefore, assigned outside the VP, to the subject. Internal arguments are all the other obligatory arguments. In this respect, the information supplied by theta-grids differs from that supplied previously by strict subcategorisation in that the subject was not subcategorised for by its verb. The asymmetry between (D-structure) subject and object is kept in that the first is indirectly theta marked via INFL whereas the second is directly theta marked by the verb.

- The order of the complements of a head is not indicated in the lexical entry of that head, but rather derived through Case theory -- Case being assigned directionally under adjacency, for example.
- The syntactic categories of the complements of a head are not specified in the lexical entries; instead a set of statements of Canonical Structure Realisation (CSR) express the relation between a particular theta role, *C*, and one or several syntactic categories, (CSR(*C*)). This notion is first proposed by Grimshaw (1981), though she does not dispense with category selection (i.e. subcategorisation) altogether; the idea is pursued by Pesetsky (1982) and others. CSRs are exemplified by the following, from Chomsky (1986a):

(2) CSR (patient)	is NP
CSR (goal)	is NP
CSR (proposition)	is NP or clause

CSR rules work together with Case theoretic information, in such a way that the category of a complement is derivable from their interaction. For instance, under the assumption that clauses are not assigned Case (Gràcia 1986:42), a verb which does not assign Case may have as complement a clause, but not an NP without a preposition.

Altogether, the three kinds of information above aim to provide without redundancy the information provided in previous models by strict subcategorisation, with some additional advantages, besides economy, the most important being the epistemological priority of Case theoretic and thematic properties (cf. Pesetsky 1982). For an exposition of the lack of epistemological priority of subcategorisation (resulting in grammatical relations, i.e. the relations of subject, object, etc.) as opposed to semantically relevant features (expressed by theta roles, such as

AGENT, THEME, BENEFACTIVE, LOCATIVE, and INSTRUMENTAL) and morphosyntactic marking (e.g. Case), see Chomsky 1982:118 ff.

All this means is that, in formulating principles of word formation, subcategorisation frames are not available, so that lexical entries like the following are not at our disposal:

- (3) a. kenkyuu 'to study' [AG[TH --]]
 b. hannyuu 'to bring in' i) [AG[TH --]]
 ii) [AG[TH[GOAL --]]]

These are Japanese examples from Shibatani and Kageyama 1988; the verb is a sister of the THEME position in the first two cases, and of the GOAL position in the third; the AGENT is in the sister position of the constituent made of THEME and verb in (3a), and so on. In consequence, an analysis on the basis of this kind of lexical entry is excluded.

As a result, lexical entries contain the following information:

- (4) xxx phonological specification
 (+/- nominal/verbal)
 (external theta role, internal theta role(s))
 [Case(s) assigned]
 lexical semantic specification¹

This is exemplified in (5), which would be typical of a rich case language (taken from Baker 1988a:112):

- (5) steal /.../
 [+V, -N]
 (AG, TH, SOURCE)
 | |
 ACC ABL
steal

Yet (5) indicates directly how particular theta roles are realised (THEME as ACC, and so on). There are other proposals, which aim at expressing generalisations in the relation Case/theta role which is explicitly encoded in (5). In these proposals, a head is marked as [+Case] or [-Case], depending on whether it does or does not

¹I shall leave unspecified how lexical semantic information should be included. It is clear, though, that this specification should tell apart two entries with identical category, thematic and Case theoretic properties if these two entries are not synonymous.

assign Case, this being a characteristic of a lexical item. (6) exemplifies a lexical entry of this type.

(6) steal /.../
 [+V, -N]
 (AG, TH, SOURCE)
 [+ Case]
steal

I shall illustrate how this approach works with Foley and Van Valin's (1984) model. This involves an actor-undergoer hierarchy for the interpretation of the arguments of a head:

(7) ACTOR: Agent
 ..
 Effector
 ..
 Locative
 ..
 Theme
 ..
 UNDERGOER: Patient

"In interpretative terms, Figure 1 [(7) above] signifies that the most natural or unmarked interpretation for the actor is agent and for the undergoer is patient, the other interpretations being progressively more marked" (Foley and Van Valin 1984:59-60). Crucially, we would have a language specific correlation between the levels in the above hierarchy and the cases which mark them overtly.

In this approach, heads have to be marked in the lexicon for the theta role they assign to their arguments and also must be specified as to whether they assign Case or not, but the actual Case they assign is specified in a general way for the language (in the correlation hierarchy/Case). This allows us to dispense with

particular linkings in lexical entries such as those in (5)².

In fact, even those who used specifications like the one in (2) admit to at least partial predictability of the relation Case/theta role; for example, in the view of Baker 1988a:455: "I assume that (apart from "quirky case") this link [between a verb's Case assignments and its theta roles] does not need to be stipulated, but follows from more general principles". Of course, the problem remains to decide what "quirky case" is, i.e. to find out if the linkings Case/theta role are subject to statements of exception.

I regard the choice between this last approach to Case assignment and the previous one (i.e. the one exemplified in (5)) as an empirical matter; that is, it is not clear that, actually, all Case assignments are predictable from a single ranking of Cases in a given language. Of the two possibilities, the most restrictive one would include only a [+Case] or [-Case] specification. Because of its superiority from the point of view of economy, this kind of lexical entry (e.g. (6)) will be used henceforth.

3.1.2. Principles of the grammar

Given the subtheories of the grammar postulated in Chomsky 1981 (see (8)), the first question to ask is: Which of them are relevant for word formation and, in particular, which play a role in compounding, more specifically compounding in the

²There is a problem in connection with the actor/undergoer hierarchy. Foley and Van Valin account for the difference between (i), (ii) and (iii) (reflecting Foley and Van Valin's grammaticality judgements) by saying that in each case the undergoer of the action expressed by the verb is interpreted as a possibly different theta role (in the terminology here).

- (i) a. Max gave flowers to the girl.
b. Max gave flowers.
c. ?Max gave the girl.
- (ii) a. John drained the water from the pool.
b. ?John drained the water.
c. John drained the pool.
- (iii) a. Fred loaded the hay.
b. Fred loaded the truck.

With *give*, the undergoer must be interpreted as the Theme, not as a Locative, so that *the girl* in (ic) cannot be interpreted as Locative, but only as Theme. *Drain* works differently: the undergoer is interpreted as a Locative, not a Theme (hence (iic) is much better than (iib)). *Load* accepts the two interpretations: (iiia) has a Theme undergoer, (iiib) a Locative undergoer. (This has other manifestations, e.g. in word formation, cf. *pooldrainers*, **waterdrainers* vs. *vanloader*, *hayloader*). The fact that, for a particular verb, a role #1 in the hierarchy outranks #2, while the reverse happens with another verb, is completely idiosyncratic and, therefore, must be stated in the verb's lexical entry. The only place to express this in the kind of lexical entries I have presented so far is the theta-grid: a head's theta-grid will contain only the obligatory theta roles (the Theme for *give*, the Locative for *drain*) or will have alternative theta-grids (a Locative or a Theme for *load*).

Romance languages? That is, which are the devices we must refer to in order to account for the existent and the possible compounds and for the impossibility (i.e. ungrammaticality) of others? Moreover, are those devices the same as those that one needs to explain other word formation phenomena or not? Notice that if the answer to this last question is negative, we will be obliged to recognise that there is no such thing as devices used for all and only word formation phenomena.

- (8) Subtheories of principles
- (i) Case theory
 - (ii) government theory
 - (iii) theta theory
 - (iv) binding theory
 - (v) bounding theory
 - (vi) control theory

3.1.2.1. Morphological versus lexical

Let us start with a terminological clarification concerning the relation between morphology and the lexicon. This is due to Dowty (1979); his approach is not a principles-and-parameters one, and so he emphasises the notions of rule and operation, which are not consistent with the framework that I adopt. However, his insight still seems useful, and I shall present it as initially formulated. (If a translation were required, rules refer to *how* a form is produced, operations to *what* is produced).

This classification is summarised in the following table. The phenomena noted in each box have to be understood as prototypical representatives of each kind of operation/rule combination.

	syntactic rule	lexical rule
syntactic operation	e.g. S generation I	idioms II
morphological operation	inflection; derivation in polysynthetic languages III	typical word formation IV

The Strong Lexicalist Hypothesis (as formulated in Lapointe 1981, Jensen and

Stong-Jensen 1984, Di Sciullo and Williams 1987, etc.) regulates precisely the relation between the lexical and the syntactic in the table above. This hypothesis has been formulated as follows by Anderson (1982), who does not subscribe to it himself.

- (9) Strong Lexicalist Hypothesis
Syntactic rules cannot make reference to any
aspects of word-internal structure.
(Anderson 1982:573)

According to a model which does not respect the Strong Lexicalist Hypothesis, words can be generated inside the lexicon or outside it -- i.e. words are formed by *lexical* rule (that is, a rule applying in the lexicon) or by *syntactic* rule. However, if one deals with words (as opposed to phrases, sentences, etc.), the *operations* required are called *morphological*, not *syntactic*. The only units included in the lexicon but which are operated on syntactically are idioms and the like (see Bar-Hillel's definition of idiom, section 1.3.).

If one admits to having morphological operations triggered by syntactic rules (the ones in III), the Strong Lexicalist Hypothesis is discarded: this hypothesis precludes any interaction between the internal structure of words and the syntax; inflected words are produced in the lexicon and the syntactic rules only have access to features expressed at the word level. III then corresponds to word formation outside the lexicon, while IV corresponds to word formation in the lexicon. The possibility of having morphological operations by syntactic rule will be considered on the basis of the data on Catalan compounding. As a result, it should become apparent in what fashion a GB grammar can discriminate a range of facts classifiable as III or IV.

3.1.2.2. GB and word formation

Of the six subtheories mentioned above, I am going to concentrate on the first four: theta theory and binding theory (in this chapter), Case theory (in Chapter Four), and government theory (in Chapter Five).

We work on the assumption that, unless it is proved necessary, Catalan compounding (and, with it, Romance compounding in general) does not involve the intervention of Move α in its generation. (This is examined in some detail in section 4.2.1.). In consequence, bounding theory does not play any role in Romance compounding, in so far as it constrains movement. Secondly, the structures treated by control theory present embedded clauses whose subject is PRO, and I assume that there is no such thing as word-internal embedded sentences, so that control theory can also be left aside for my purposes. Notice, however, that I do not

take for granted that coindexing relations are not going to be found word-internally, though this is an assumption underlying the Lexicalist Hypothesis. Thus I do not rule out the possibility of binding theory playing a role in word formation.

As a consequence of the fact that Move α is not used for Romance compounding, one can deduce that, among other principles, the Empty Category Principle (ECP) plays no role at all here, unlike in Baker's account of incorporation (cf. Baker 1988a), which involves X^0 movement. Given that the ECP states that traces must be properly governed, and traces are left behind by the application of Move α , if we do not have movement, the principle cannot apply. Only a Generalised Empty Category Principle (GECP; cf. Chomsky 1981:274) could possibly apply to empty categories which do not result from movement, but are base generated.

There is another fundamental principle of GB, the Projection Principle (PrPr), which states that lexical features (such as Case assigning properties, thematic properties and syntactic category as projected from the lexicon) must be represented at every syntactic level. If there are to be operations altering the thematic structure of a lexical item in the grammar, because of the PrPr, the operation is excluded from the syntax and it is, thus, bound to be in the lexicon. This idea is the basis of Borer's (1984) work. Because there are rules which violate the PrPr and so must be presyntactic, she derives part of the content of the Lexicalist Hypothesis from the PrPr, while the rest of it is considered a spurious generalisation. In Borer's analysis what is traditionally called inflection is allowed to happen outside the lexicon since it does not violate the PrPr; other rules (the "derivational" ones) occur presyntactically because they do violate the PrPr³. There is a sense in which it is the main aim of this thesis to answer the question: Does compounding fall within the scope of the PrPr or not?

3.2. On some universal constraints of morphology

Shibatani and Kageyama (1988) give a summary of the properties that words have, whether they are formed in the lexicon or postlexically, i.e. outside the lexicon (these authors do not assume the Lexicalist Hypothesis, but rather argue against it). These properties have been expressed in linguistic theory by a series of disconnected principles. I briefly review them here.

1. Case particles can be absent from nouns, etc., once they are part of a compound.

³A topic for future research is: Which alterations of lexical features by lexical rules are allowed? They cannot *all* be allowed, but, still, which are the ones allowed?

2. Tense is excluded from the verbal elements. The authors invoke "the general principle banning tense inflection in word formation" (op. cit.:462).
3. Words constitute morphological units that "resist syntactic interruption" (op. cit.:462).
4. The only structure available is binary. This was expressed by Selkirk's (1982) Binary Branching Condition.
5. "A transitive verb is allowed to combine only with its first sister noun (or, in GB terminology, the noun that is properly governed or 'L-marked') in the verb's (extended) strict subcategorization frame" (op. cit.:463). This was first formulated by Roeper and Siegel (1978) as the First Sister Principle. Later, Mithun (1984) has noted the impossibility of (noun-)incorporation of transitive subjects. Selkirk's (1982) First Order Projection Condition (which states that "all non-SUBJ arguments of a lexical category X_i must be satisfied within the first order projection of X_i " 1982:37) is dismissed by the authors in the light of Japanese compounds. This area of study has developed considerably in recent years; our treatment will rest on notions other than subcategorisation frames and grammatical relations, given the restrictions on lexical entries expressed above.
6. Word formation is lexically governed -- leaving room for idiosyncrasy in the statement of rules⁴. I dismiss this point in the light of what has been discussed in 1.3. As Baker (1989) has put it in his analysis of serial verb constructions (SVC): "I take these [examples] to be SVC idioms, with essentially the same structure as [regular SVC]. They do not imply that SVCs are lexically formed, just as *kick the bucket* and *pay heed* do not imply that English verb phrases are lexically formed" (Baker 1989:535, note 15).

On the basis of Japanese compounding again, Shibatani and Kageyama characterise lexical word formation with respect to postlexical word formation.

7. Word formation rules apply to X^0 categories, i.e. the constituents of words are, at most, other words -- not phrases. This constitutes what Botha (1981) called the No Phrase Constraint, formulated already in Roeper and Siegel 1978. This constraint is loosened for postlexical compounds (op. cit.: 471-472). In fact, it may turn out to be too

⁴However, postlexical compounds (i.e. compounds formed in the syntax) are semantically transparent (compositional in meaning), in Japanese.

strong a constraint even for lexical compounding, since one can find (presumably lexical) compounds like *People With Aids Alliance* (op. cit.:472) or, by the same token, *No Phrase Constraint*, where *people with aids* and *no phrase* appear to be phrases.

8. Part of a lexical word cannot hold an anaphoric relationship with another item elsewhere, while this seems allowed for postlexical words. The constraint on lexical words is known as the Anaphoric Island Constraint (cf. Postal 1969), restated in Williams 1981b as the Atom Condition.
9. "Lexical words in general cannot participate in anaphoric relations because they are formed in the lexicon without any reference to syntactic environments, and because their primary function is naming, where only generic expressions without referential function come in" (op cit:476). This lack of referentiality or, rather, generic quality was already noticed by Bally (1950), and contrasts with the referentiality attributed to the constituents of postlexical compounds by Shibatani and Kageyama.

[1-6] and [7-9] include a large part of the principles of the theory of morphology of recent years.

Of these principles, some can be reformulated and naturally be subsumed by a subtheory: e.g. [1] to Case theory, and [5] to an (extended) theta theory. Most of the other statements can be considered definitional: they follow from the notion of lexicon or should follow from a theoretical conception of complex word. Most of them will be of recurrent interest.

3.3. Subtheories related to meaning

The purpose of this section is to examine, in a GB framework, semantic aspects of Catalan compounds.

3.3.1. Basic assumptions

In view of examples such as (10), (11) and (12), one might wonder whether it is possible to predict the meaning of compounds, or whether alternatively their meanings are only established by use, if established at all.

- (10) *figa-flor*
fig flower 'weak character'
- (11) *&porta-barrets*
carry hats 'somebody/box/etc. that carries hats'
- (12) *ala-trencar*
wing break-INF 'to break some-one/thing's wing/s'

Thus, the meaning of the compound in (10) does not seem to be derived in any way from the meaning of its parts; the interpretation of *figaflor* as 'weak character' escapes compositionality⁵. Only its recording would allow speakers to know what this compound means. The example in (11) is less opaque than that in (10), but, still, the form of the compound does not allow the speaker/hearer to pin down the actual content of the compound. Being a newly coined word, for which no standard interpretation is given, several uses (meanings) are possible. A Catalan speaker can, for instance, paraphrase (11) as 'somebody or something which transports hats', so that the meaning of the compound is so vague that it can be used to refer, in context, to things as diverse as 'person whose job is to transport hats' 'box to carry hats' or 'van designed to transport hats'. Finally, the example in (12), which is an existing compound, has a specific meaning, namely 'to break some-one/thing's wing/s'; there are other compounds of Catalan which are interpreted in the same way, e.g. *cortrencar* 'to break someone's heart'. It does not seem that, in these cases, other interpretations are available.

Examples (10) through (12) suggest that there are various degrees of predictability in the meaning of compounds, or, in other words, varying degrees of semantic transparency. While (10) is totally opaque, (12) seems to be transparent, and the example in (11) is semi-transparent (or perhaps transparent but vague).

Given this state of affairs, the question is whether it is possible to account for the meaning of all the compounds in a similar fashion. If so, given that (10) is totally unpredictable in meaning, one would have to conclude that the meaning of compounds is unpredictable in general. The meaning of the existent compounds would be listed in the lexicon and new compounds would be coined and at the same time attributed meanings on a completely idiosyncratic basis. Adopting this position would amount to abandoning the idea of making any generalisations on meaning, and to establishing regularities at other levels only (morphological regularities, for example). However, the fact that the interpretation of new compounds such as (11) is not completely free argues against a uniform semantic treatment of compounds; i.e. their meaning is not arbitrary. Their pattern of interpretation can minimally be restricted as follows:

⁵See footnote 1, Chapter One.

"A novel compound $\alpha\beta$ denotes some set (exactly which one we do not know) such that members of this set are β 's and are typically associated by some appropriately classificatory relation to an α . Needless to say, this represents only a rough approximation of a compounding rule (though in itself it is precise, relative to the two constants)." (Dowty 1979:319)

Is it possible to interpret (12) by this same principle? The example in (12) has a more restricted interpretation than that in (11); but perhaps this in virtue of it being listed (i.e. in virtue of it having been used in a particular way and repeatedly). The possibility exists, however, for the creation of new compounds, such as *dittrencar* 'to finger-break', whose meaning would be 'to break somebody's finger/s', and whose interpretation would not vary according to context -- hence it would not be interpreted as 'to break with the finger', or 'to break when looking at a finger'. Then, if we decide that it would only be grammatical to express, by the compound *dittrencar*, the notion 'to break somebody's finger/s', our grammar (and not pragmatic constraints) must attribute to this compound the only possible interpretation it can have.

In consequence, a distinction should be drawn between (10), (11) and (12), if we are to account for the interpretations that native speakers give to them. My assumption is that the meaning of the compounds in (12) is not vague and only pragmatically restricted, as expressed in Dowty's interpretation principle (which accommodates (11)), but rather, grammatically determined. As for (10), it cannot be assigned meaning according to a general pattern, and its interpretation falls within the domain of the idiosyncratic.

Having settled this point, the next question is: By what means are the semantic properties of grammatical structures characterised in a Government and Binding model? Theta theory is the subtheory that deals with the semantic relations between a lexical head and its complements. I shall entertain the hypothesis that the relationship between a head and an argument can be held word-internally.

The first attempts to relate nouns which are part of compounds to theta-marked arguments in the literature on Catalan appear in Mascaró 1986 and Cabré and Rigau 1986. These works examine the compounds in (13), as reported in 2.8.

- (13) a. *cama-trencar*
 leg break-INF 'to break someone's leg/s'
- b. *para-brises*
 stop breezes 'wind screen'

The analysis here will differ in quite a few respects from theirs. The hypothesis underlying my analysis is that the lexical properties of the words at the base of a compound are identical to those of the corresponding simple words. As part of their lexical entries, *trencar* and *parar* contain the theta-grids in (14).

- (14) *trencar* [AG, TH]
parar [AG, TH]

Of the theta roles which these two verbs assign, *cama* 'leg' in *camatrencar* and *brises* 'breezes' in *parabrisas* correspond to the THEME of the verb, i.e. the entity which is broken in (13a) and the entity which is stopped in (13b). Also, in compounds of the *parabrisas* type, the compound as a whole designates either the AGENT, or the INSTRUMENTAL or the LOCATIVE of the predicate:

- (15) a. *guarda-roba* *guardar* [AG, TH, (LOC)]
 keep clothes
 'cloak-room' (LOC)
 & 'somebody who keeps clothes' (AG)
- b. *obre-llaunes* *obrir* [AG, TH]
 open tins
 'tin-opener' (INSTR)

Mascaró (1986) gives a list of compounds belonging to these two types where the noun does not appear to correspond to the THEME of the verb:

- (16) a. *gira-sol*
 turn sun 'sun-flower'
- plora-miques*
 cry littles 'cry-baby'
- prega-déu*
 pray god 'praying mantis'
- b. *cor-nuar*
 heart knot-INF 'distress'
- peu-calcigar*
 foot step-on-INF 'to step on somebody's foot'
 or 'to step on with one's foot'
- c. *torna-vís*
 return screw 'screw driver'
- marxa-peu*
 go foot 'doorstep'
- d. *salta-martí* 'tumbler'
 jump martin
- caga-tió* (Christmas log)
 shit log

Some of these will be accounted for in 3.2.2.1. and 4.2. For the moment, I shall just give an informal classification of the compounds in (16a-d). In (16a) and (16b), taking the noun to be the THEME presents some difficulties; for instance, in

pregadéu, déu does not seem to correspond to the THEME of *pregar* 'to pray', but rather the GOAL. The examples in (16c) conform to the morphology of Catalan compounds, but are in fact gallicisms, from the French *tournevis* and *marchépied* (where *tourner* 'to turn' and *marcher* 'to walk' do not mean the same as *tornar* and *marxar*, despite the common etymology, and have different theta-grids). In (16d), the compound derives from a sequence of an imperative and a vocative; for instance, *cagatió* designates a piece of wood to which children sing a Christmas song saying 'shit, log...', to get presents, etc. These expressions are morphologically identical to *parabrises*, but their relation to a sentence in the imperative mood is fairly transparent for a native speaker. Of course, the examples like those in (15) are not amenable to this interpretation: *guarda-roba* 'cloak-room' cannot mean, on any account, 'tidy up, clothes!', i.e. it is not a noun derived from an order to tidy up addressed to some clothes.

Supposing, then, that the analysis in terms of argument/predicate is possible for the compounds in (13), what other relations of this kind can one expect to find within compounds, given the basic lines of theta theory?

Theta theory establishes that only some categories can be in argument position: nominals, and sentences (which are irrelevant here). Also, typically, verbs and prepositions are argument-taking lexical items; though some adjectives and nouns may have argumental structure, most of them do not require complements and, therefore, do not have theta-grids.

There are sixteen possible binary combinations of the four major categories. If argument/head relations occur typically with a verb or a preposition, there is nothing to be said about the meaning of [N N], [N A], [A N] and [A A] compounds (with the exception of a few cases of theta role assigning adjectives). For the rest, it would be desirable to account on the basis of theta theory for the existing and the possible compounds of Catalan, and the gaps left in the lexicon.

Lieber 1983 is a specific proposal to account for the existing and the possible compounds of English in GB; she postulates the Argument-linking Principle (Lieber 1983:258), whereby a word-internal argument-taking category must be able to satisfy its argumental requirements; the first clause of this principle follows from the PrPr when compounding occurs in the syntax.

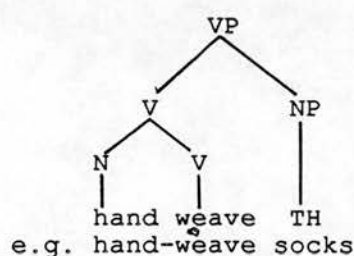
(17) Argument-linking Principle

- a. In the configuration $[]_{V/P} []_{\alpha}$ or $[]_{\alpha} []_{V/P}$ where α ranges over all categories, V/P must be able to link all internal arguments.
- b. If a stem $[]_{\alpha}$ is free in a compound which also contains an argument-taking stem, α must be interpretable as a semantic argument of an argument-taking stem, i.e. as a Locative, Manner, Agentive, Instrumental, or Benefactive argument.

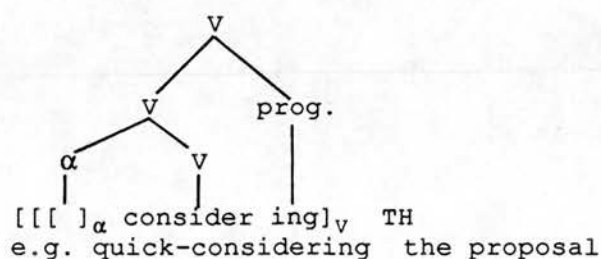
This principle only applies to compounds, which are the only words with the configuration presented in (17a), since, if α was an affix, the representation would be $[[]_{\alpha}]$ or $[\alpha []]$, since an affix has no word bracket. Lieber's conception of internal argument is standard; semantic arguments are "phrases which are not obligatory or lexically specified"; finally, a stem is free "if it is left unlinked by an argument-taking lexical item" (Lieber 1983:257).

The interaction of the Argument-linking Principle with Lieber's percolation conventions, whereby the righthand element in a compound is the head and shares its (especially thematic) features with the mother, is exemplified in (18) to (20)⁶.

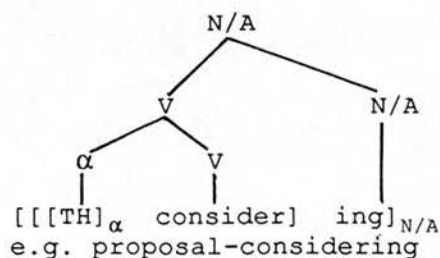
(18)



(19)



(20)

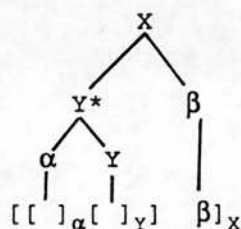


Notice that, despite Lieber's claim that primary and synthetic (i.e. deverbal)

⁶Lieber's view on these matters has changed, and in Lieber 1989 she does not consider theta-grids to be features of an argument-taking item.

compounds obey the same principles, the assignment of internal theta roles works differently in the two cases. While in primary compounds the thematic properties of the head percolate (with the other head features) to the compound node, and so theta roles are assigned outside the compound (as in (18) and (19)), in synthetic compounds theta roles can be assigned inside the word (as in (20)). Lieber assumes that features on a node of one category cannot percolate to a node of a different category, thus in a configuration such as (21), where X and Y* belong to different categories, Y must satisfy its argument in α , since it cannot link it outside Y*; not fulfilling the argument would result in a violation of the Argument-linking Principle.

(21)



However, the above mentioned feature percolation convention seems to have been overridden in that in (21) Y* does not have the thematic properties of Y. Thus it seems that thematic properties cannot be considered head features in the way that syntactic category features are. Feature percolation is further discussed in section 4.4.2.

The principles proposed by Lieber on the basis of a series of assumptions (such as binary branching, the closed-class character of prepositions, etc.) allow her to predict the possibility or impossibility of certain compounds. Moreover, they allow her to predict also their relative likelihood, since, in her own words, "a compound type containing an argument-taking stem will never be as productive as compound types [such as NN, NA, AN, AA] containing no argument-taking stems" (Lieber 1983:265); she also predicts what sorts should be nonexistent or less than fully productive (PA, NV, AV, VV, PV, VN, VA, PN). Though her predictions are largely correct, her analysis has some shortcomings which I shall consider next.

In the first place, against the feature percolation conventions, there is a group of expressions such as *call up*, *put down* and the like which consist of a verb followed by a particle and which are verbs themselves (instead of particles as predicted); the same holds for *whiteout*, which is a noun and not a preposition. Since Lieber assumes that when two stems are sisters, they form a compound, and states that compounds are right-headed in English, it would follow that a left-headed construction cannot be a compound. Thus, *call up*, *put down*, etc., if taken to be two sister stems (and Lieber does not deny this), falsify the percolation convention. To make the percolation conventions compatible with the data, *call up* and the other head initial constructions must be analysed as something other than

compounds; the same applies to exocentric constructions, such as *speakeasy*; that most of them are semantically odd (as Lieber considers them to be) does not make Lieber's analysis less inconsistent if they are included among productively derived compounds.

Secondly, it is not clear that the kind of lexical information that Lieber uses to make her predictions is available, at least not in the type of lexical entries which I assume here. Lieber rules out certain types of compounds on the basis of the specific category that an argument-taking category selects (e.g. a preposition or a verb can be specified as having an internal noun or adjective argument), but reference to specific syntactic categories is not included in the lexical entries currently assumed.

Thirdly, an attempt to rule out all impossible compounds while generating the possible ones on the grounds of theta role assignment would presuppose that the only possible grammatical relation between two elements is that between a theta role assigner and a theta role assigned constituent. This is not the case in GB, and Lieber is well aware of that when she points out that the commonest relation between two stems is between non-argument-taking stems. However, certain combinations of two categories are excluded only on the grounds that one cannot be the argument of the other (e.g. the case of [V V] compounds, Lieber 1983:265). For those cases, it would seem necessary to exclude as well other possible relationships, namely a modifier/modified category relationship.

In consequence, to overcome the problems of Lieber's account, it is necessary to provide: (i) for English, an analysis of *call up* and the like with compatible percolation conventions; (ii) a general characterisation of arguments and predicates -- that is, a relation between categories and the functions which theta theory deals with, plus an outline of the relations possible between two elements, head/complement or otherwise; (iii) percolation principles to determine the percolation of thematic properties (i.e. a principle settling whether these are head features or not, whether they are language-specific, etc.). What is necessary to make Lieber's approach sufficiently explicit is to specify lexical categories (as classes) with respect to theta theory, and to specify the percolation of thematic properties in a tree, particularly in a word structure (i.e. a tree with an X^0 mother node).

I shall deal with exocentric compounds in section 3.3.2.2. It is beyond the scope of this thesis to analyse all compounding in English (and with it the *call up* expressions). In what follows, I shall address the second issue above, namely what broad semantic relations are possible between lexical categories and how they are expressed in GB.

3.3.1.1. Theta theory

The relation between a lexical head and a complement is typically expressed by theta role assignment, governed by the Theta Criterion:

- (22) Theta Criterion
Every form at LF that requires a theta role (each argument) is associated with one and only one position to which theta roles are assigned, and each theta role determined by the lexical properties of a head is uniquely associated with one and only one argument.
(Chomsky 1981)

Under the constraints of X-bar theory, a complement is an XP (a phrasal category) of a head category X. If a theta role assignment is to take place word internally, given the assumption that generally no phrases occur inside words, complements will not be maximal projections.

The relation between theta role assigners and theta role assigned categories is exemplified by a verb and its internal arguments. Non semantically vacuous prepositions can also assign theta roles to their complements. The situation is far less clear with the other lexical categories. It is often assumed that there are adjectives which are theta role assigners, for instance *worth* (*worth a dime*) and *fond* (*fond of gin*) (from Lieber 1983:257), though apparently this is the case with only a few adjectives. As for nouns, Chomsky (1986a), for instance, assumes that they can assign theta roles to their complements (this suggestion is taken up in Gràcia's 1986 thesis on theta theory), although the issue remains open for external arguments. While (23a) is ruled out because the external argument of the verb has not been realised, i.e. because of a violation of the Projection Principle, (23b) is grammatical without the AGENT being expressed (these examples are taken from Chomsky 1986a:138).

- (23) a. *Offered a loan.
b. the offer of a loan

It is not uncontroversial to claim that, if the noun has a subject, this must be theta-marked. In fact, quite a range of functions seems to be possible for the subject of a noun phrase (cf. Chomsky 1986a:143ff). In the case of nouns without complements, in particular, the fact that expletives are excluded from subject position (as in (24)) does not prove that, if there is a subject, it must be theta-marked; in fact, the expletive is excluded anyway by binding theory (it does not belong to a CHAIN, so it is not be linked). (This example is taken from Chomsky 1986a:142).

(24) *there's book

3.3.1.2. Predication theory

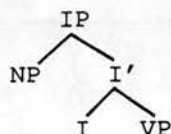
In general, predication theory deals with the relations between a predicate and its subject. This has been outlined in Williams 1980 and Rothstein 1983, for example, before the *Barriers* version of Government and Binding. In these theories, a predicate is a maximal projection which does not receive a theta role; the subject associated with it is also a maximal projection; and, finally, predicate and subject c-command each other. This can be summarised in (25).

- (25) XP is a predicate for YP
iff XP and YP stand in a relation of mutual c-command,
and
XP does not receive a theta role.

Nevertheless, applying this definition of predication to the *Barriers*' phrase structure presents a number of problems, the solution to which necessarily leads to a redefinition of predication. Moreover, the redefinition of predication here will be worked out from a slightly different perspective, closer in spirit to the model theoretic notions of (relational) functions and arguments (cf. Dowty et al. 1981).

In the first place, mutual c-command does not hold between the VP and the subject in the phrase structure in (26), given the definition of c-command in (27).

(26)



- (27) α c-commands β
iff α does not dominate β and
every γ that dominates α dominates β .
(Chomsky 1986b:8)

The subject NP c-commands the VP, but the NP is not c-commanded by the VP. If we want to consider the VP to be the predicate of the NP, two possibilities suggest themselves. M-command, as defined in (28), could be substituted for c-command:

- (28) α m-commands β
iff α does not dominate β and
every γ that dominates α dominates β ,
where γ is restricted to maximal projections.
(Chomsky 1986b:8)

Alternatively, the condition on predication could be weakened from mutual c-

command to minimal c-command of the predicate by its subject. In both circumstances, predication holds between NP and VP, as desired.

Before I go on to discuss these possibilities, another objection to the above notion of predication can be made. According to Chomsky, VP can be theta-marked by I, though this theta-marking is optional ("VP may but need not be theta marked by I" Chomsky 1986b:95, note 54). The motivation for the theta-marking of VP by I is given by the configuration in (29) (from Chomsky 1986b:69).

(29) [_I V_I [_{VP} t ...]]

(29) is the result of the movement of V to the head position of IP, I, where V and I amalgamate to V_I (i.e. an inflected verbal form). However, unless VP were L-marked, it would be a barrier, which would block antecedent government of the trace *t* by V_I. Since the movement is legitimate, Chomsky argues that V_I L-marks VP; I, not being a lexical category, cannot L-mark VP, but can theta-mark it⁷.

According to the second clause of (25), VP cannot be a predicate for the subject if it ever receives a theta role. Again, if we want VP to be the predicate of NP, a reformulation of (25) is imposed on us. A restatement which seems plausible would define predication with respect to the argument instead of the predicate; i.e. instead of saying that the predicate can never be assigned a theta role, thus that it is, possibly, a theta role assigner, one can rather state that the argument (position) is assigned a theta role. For instance, the subject may be assigned a theta role by VP, and that is what signals it as the argument of VP. (Expletive subjects will be considered below).

The relation between a VP and its subject, in a *Barriers* basic structure, is generally considered here to exemplify the predicate/argument relationship. In the configuration which motivates the theta-marking of VP, namely the raising of V to I to form an inflected verb, the subject should still be the argument of VP. However, applying the theory of predication of Williams (1980) and Rothstein (1983) would suggest that predication takes place between a subject NP and I'.

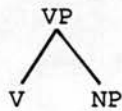
Here, a much more general notion of predication will be used. All theta role assignments are to be taken to be predicate/argument relations. An object is an argument for the verb which assigns a theta role to it, and the object of a preposition is the argument of the preposition. In (30), the object *xocolata negra* 'plain chocolate' is the argument of *menjar* 'to eat', and *la terrassa* 'the terrace' is the argument of *a* 'on'.

⁷For an alternative view, see Cann and Tait 1990.

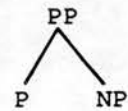
- (30) a. vull [menjar_V [xocolata negra]_{NP}]
 I-want eat-INF chocolate black
 'I want to eat plain chocolate'
- b. llegeixo [a_P [la terrassa]_{NP}]
 I-read on the terrace

We require a modification of the definition of predication, given that the X-bar constraints in (25) conflict with the situation in (30). In (17), both predicate and argument are phrasal categories; in (30), the predicates are X^0 categories, a verb in (30a)-(31a), and a preposition in (30b)-(31b).

(31) a.



b.

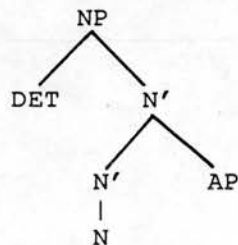


Other semantic relations which will be considered predications hold between an adjective and the nominal it modifies, and an adverb and the VP it modifies.

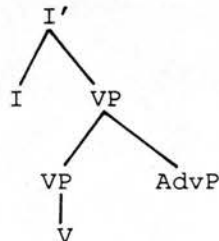
- (32) a. Veig una casa blanca.
 I-see a house white
 'I see a white house'
- b. Llegeix lentament.
 reads slowly

In (32a) it is predicated of the house that it is white; and in (32b), it is predicated of the reading that it is slow. N' and VP, in (32a) and (32b) respectively, are arguments of an AP and an AdvP. Notice that in (33b) VP is the argument of the AdvP because the AdvP is not theta-marked, while the VP is.

(33) a.



b.

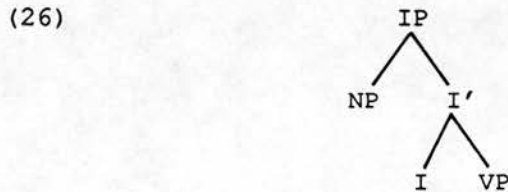


If the adjectival phrase is the predicate of N' , we must conclude that arguments are not necessarily phrasal (i.e. XP). I shall reformulate (25) in accordance with this and, also, the conclusion above that predicates are not necessarily phrasal, either.

So, in the preliminary reformulation of (25), (34) below, X and Y stand for any projection of a category.

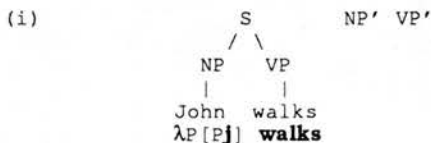
- (34) X is a predicate for Y
iff X and Y m-command each other, and
Y can be theta-marked.

(34) appeals to the notion of m-command, and therefore represents a weak version of predication when compared to one involving c-command. The question remains whether mutual m-command or c-command of the predicate by the argument is preferable. Aoun and Sportiche (1983), for instance, have argued for the notion of m-command to take priority over the notion of c-command in the formulation of the theory of government. As for my formulation of predication, the decision must be taken on the basis of the structures in which predicate/argument relations hold in principle, that is, the ones displayed in (26), (31) and (33) above. For our purposes, evidence can be found against a definition in terms of m-command such as (34). In (26), repeated for convenience, subject and VP m-command each other and are both possibly theta-marked, NP by VP, and VP by I.

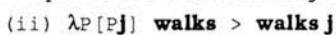


Hence, in (26), if the definition (34) is adopted, the subject is the argument of the predicate VP, and VP is at the same time the argument of the predicate NP -- an undesirable result⁸. To make sure that the subject is the argument of VP, and not viceversa, a definition in terms of c-command is chosen, since the subject c-commands the VP but not viceversa. This move will have no undesirable effect in determining arguments and predicates in (31) and (33), because in all cases the argument c-commands its predicate. Thus,

⁸It is possible to "type-raise" the lexical semantics of subjects so that subjects are uniformly treated as functions over VP in the phrasal analysis (cf. Montague 1974 [PTQ]).

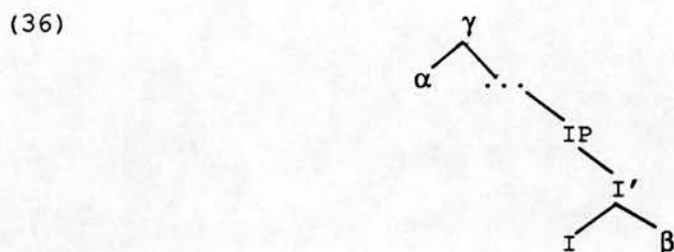


However, the result of simplifying the phrasal and lexical semantics shows that the effect is still to treat verbs as predicates over their subjects:



- (35) X is a predicate for Y
iff Y c-commands X,
X m-commands Y, and
Y can be theta-marked.

As a result, we appropriately avoid having a constituent predicating an argument in a higher clause, as in (36).



Also, it must be stated that predication as in (35) is established at D-structure and holds after it; i.e. it is transmitted through a chain. In this way, the following situation is avoided: in (37), the verb trace would be a predicate for the complement NP. Traces are not arguments or predicates, only the heads of their chains are.



A consequence of such an approach to predication is that expletives are not arguments because they cannot be assigned theta roles (for another argument leading to the same conclusion, see Cann and Tait 1989). Note, however, that they are arguments for Williams, Rothstein and the authors who follow them, e.g. Baker (1988a). Therefore, weather verbs, for instance, are zero-place predicates, whether an expletive must appear in the surface or not -- compare (38a), from Catalan, and (38b), from English.

- (38) a. Plou.
rains
b. It rains.
*Rains.

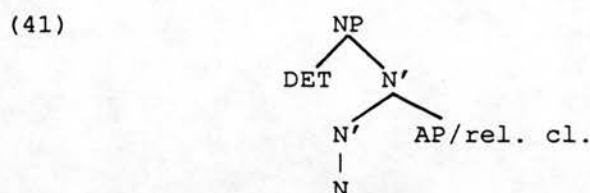
One of the motivations for the reformulation of (25) here has been that VP is theta-marked by I. Nevertheless, Chomsky points out that the verb daughter of VP cannot be theta-marked by I, because if it were this verb would be properly governed and, therefore, long distance movement would be permitted, e.g. in (39) (from Chomsky 1986b:68). Because (39) is ungrammatical, movement has to be prevented in these circumstances.

(39) [how tall]_j be_i [_{IP} John [_{IP} will [_{VP} t_i t_j]]]

So, theta-marking of VP (by I) cannot percolate down to V. To prevent it from percolating, the qualification (40') on theta-government is abandoned by Chomsky (1986b:19).

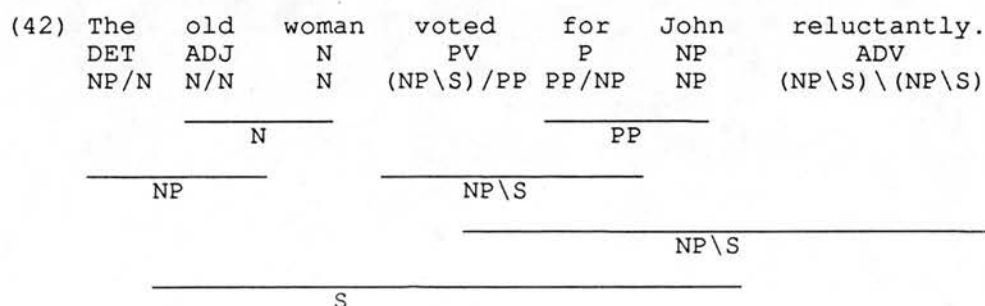
- (40) α θ -governs β
 iff α is a zero level category that theta-marks β ,
 and α , β are sisters
 (40') or β is the head of a sister of α .

However, it is necessary to allow theta-marking to percolate down in other circumstances, e.g. (41), if the adjective is to predicate of the N'.



Despite the fact that in Chomsky 1986b (40') is considered to have undesirable effects (to amount to the undesired theta-marking of V), it will be assumed that (40') will be kept for L-marking; thus α theta-governs β if and only if (i) α is a zero level category that theta-marks β , and α , β are sisters, or (ii) β is the head of a sister of α , and α L-marks β . As a consequence, I -- as a non-lexical category -- has the ability to theta-mark its sister VP only, unlike lexical categories, which theta-mark their sisters and the heads of their sisters. This idiosyncrasy corroborates that the I-projection system is defective, as it has been argued to be with respect to barrierhood (cf. Chomsky 1986b).

In conclusion, (35) makes it possible to keep together under the term of predication all theta role assignments, plus the relations of modification of nominals and verbs by adjectives and adverbials. It is worth remarking that this corresponds closely to functor/argument relationships in categorial grammar; predicates here correlate with functors in CG, as illustrated in (42).



Consequently, the definition of predication in (35) can be seen as a basis for a translation mechanism between, on the one hand, functors and arguments in CG and, on the other hand, predicates and arguments at D-structure in GB.

It might be argued that the theory of predication outlined here differs in essence from that of Williams, Rothstein and others, in that it is conceived as including theta theory, instead of being complementary with it. Therefore, it has shifted in emphasis to comprehend semantic relations which were not included originally. Yet, in its initial formulation, the original predication theory is not capable of handling that for which it was designed (the sentential subject/VP relation), thus a reformulation is, in any event, required.

In the remainder of this chapter I shall analyse some compounds of Catalan displaying predicate/argument relations; one subsection is devoted to the *alatrençar* 'to wing-break' type, another to the *parallamps*, 'stop-lightning', 'lightning conductor' type. The first type is related to inalienable possession, for which an analysis is required. The second type exemplifies exocentric compounds, and so the analysis will involve a treatment of exocentricity extendable to other exocentric constructions. Both compound types display theta role assignment, i.e. a particular kind of predicate/argument relationship; in Chapter Six, some other compounds will be examined which display other kinds of predicate/argument relations.

3.3.2. *Compounding and theta theory*

3.3.2.1. *The alatrençar type*

The central property of this type of compound, and one which has been remarked by many grammarians, is that the noun in the compound generally refers to a part of the body and is in a relation of inalienability with its possessor (the term "possessor" is to be understood colloquially; it is not a technical term). Surveys of the expression of inalienability in Catalan are found in Argente 1975, Bartra 1985 and Bonet and Solà 1986. However, none of these works accounts for inalienability in compounding and the specific problems that it represents.

Consider the grammaticality of the compounds in (43), in contrast with the ungrammaticality of those in (44), where the noun does not correspond to an inalienable part of the body.

- (43) ala-trencar
wing break-INF 'to break some-one/
thing's wing/s'
- coll-tòrcer
neck twist-INF 'to twist someone's neck'
- sang-glaçar
blood freeze-INF 'to paralyse'
- (44) *gerra-trencar
jar break-INF
- *llapis-tòrcer
pencil twist-INF
- *vodka-glaçar
vodka freeze-INF

This is so given the ontology of inalienability for this world; (44) need not be ungrammatical in a world where e.g. vodka were an inalienable part of the body. What varies across worlds is what is inalienable, not the grammar itself. The compounds in (43) are not the only compounds in Catalan in which the notion of inalienability plays a role. In most compounds of the form [N A]_A, the noun denotes an inalienable part of the body, as shown in (45), to be compared with (46).

- (45) cama-curt
leg short 'short legged'
- cua-llarg
tail long 'long tailed'
- panxa-content
belly happy 'laid back'
- (46) *faldilla-curt
skirt short
- *cotxe-llarg
car long
- *gos-content
dog happy

The data above shows that inalienability is a concept that has repercussions in the grammar, relevant for syntax and semantics. As a grammaticalised notion, the expression of inalienability varies cross-linguistically. For an instance of the difference between English and Catalan, see (47) and (48), where IA stands for "inalienable possession" interpretation, A for "alienable possession" and noP for "no possession".

- | | | |
|------|--------------------------------|----------|
| (47) | John cleaned his teeth. | IA or A |
| | John cleaned the teeth. | A or noP |
| (48) | El Joan es renta les dents. | IA |
| | DET Joan SE cleans the teeth | |
| | El Joan renta les dents. | A or noP |
| | DET Joan cleans the teeth | |
| | El Joan renta les seves dents. | A |
| | DET Joan cleans the-his teeth | |

Catalan seems to behave very much in the same way as other Romance languages such as French with respect to the absence of possessive pronouns in expressions of inalienability (cf. Guéron 1984, where data from French is compared with data from English).

A description of the compounds in (43) is given in (49).

- (49) [N V]_v
 where
 (i) N is linked to the TH in the V's theta-grid,
 and
 (ii) N is an inalienable part of the body.

(i) rules out, for instance, *ull-mirar* 'eye-look at' meaning 'to look with one's eye/s', i.e. with *ull* 'eye' as an INSTRUMENTAL, and (ii) excludes (44) (e.g. **gerra-trencar* 'to break a jar').

Consider (50). In (50b-c) the sentence with *alatrencar* (50a) is paraphrased without the compound.

- (50) a. El caçador alatrenca l'ocell.
 the hunter wing-breaks the bird
 'The hunter breaks the birds's wing/s'
- b. El caçador li trenca l'ala (a l'ocell).
 the hunter DAT breaks the wing (to the bird)
 'The hunter breaks the bird's wing'
- c. El caçador li trenca les ales (a l'ocell).
 the hunter DAT breaks the-PL wings (to the bird)
 'The hunter breaks the bird's wings'
- d. *El caçador trenca l'ala a l'ocell.
 the hunter breaks the wing to the bird
- e. El nen li trenca la joguina (a la seva germana).
 the boy DAT breaks the toy (to his sister)
 'The boy breaks his sister's toy'

- f. El nen trenca la joguina a la seva germana.
the boy breaks the toy to his sister
(same meaning)
- g. *El caçador li alatrenca l'ocell.
the hunter DAT wing-breaks the bird
- h. *El caçador l'alatrenca l'ocell.
the hunter it wing-breaks the bird
- i. El nen trenca la joguina.
the boy breaks the toy
- j. El caçador trenca l'ala.
the hunter breaks the wing
(i) ok
(ii) * if IA
- k. *El caçador alatrenca.
the hunter wing-breaks

What is shown in (50b-d) is that the dative clitic *li* 'to him/her' is obligatory when the referents of the direct object and the indirect object of the verb are related by inalienability; thus (50d) is out because the clitic is absent. This contrasts with (50e-f) where the relation is not one of inalienability: *la joguina* 'the toy' is not an inalienable part of *la germana* 'the sister'. (50e) shows that the clitic *li* can cooccur with the full BENEFACTIVE NP, and (50f) indicates that it need not. So, despite the superficial resemblance of (50c) and (50e), at some level of representation the two sentences must be different.

On the other hand, the sentences where the compound occurs do not allow a coreferential dative clitic *li* (as in (50g)⁹), nor a coreferential accusative clitic *el/l'* (as in (50h)) -- which matches with *l'ocell* in Case; so, while (50g) could be ruled out on the basis of a mismatch of Case between dative *li* and accusative *l'ocell*, this explanation is not available for (50h). (50h) can be ruled out because it is true of sentences with and without compounds that copresence of an accusative clitic and a coreferential full NP (ACC clitic doubling) is ungrammatical in Catalan (e.g. **El veig el llibre* 'I see it the book'). So, it is possible to exclude (50g-h) on the basis of Case concord and considerations of clitic doubling in Catalan. Overall, in the sentences without the compound, the clitic is obligatory for the expression of inalienability; in sentences with the compound accusative and dative clitics are forbidden to refer to the possessor.

The last three sentences illustrate that it is obligatory for the possessor NP to surface in the sentence if the relation between possessor and possessed is one of

⁹(50g) is grammatical with a meaning other than the intended one: 'The hunter wing-breaks the bird to him/her'. Why this reading is possible will become clear as I proceed.

inalienability, but not otherwise. Thus (50i) is grammatical where no possessor of *la joguina* 'the toy' is expressed; by contrast, (50k) is ungrammatical because the possessor, i.e. *l'ocell*, is not expressed and in the compound *ala* is necessarily an inalienable part. (50j) shows clearly that the possessor is only required for an inalienably possessed referent, so there is a reading for (50j) without the possessor: 'the hunter breaks the wing', but a wing which does not belong inalienably to anybody.

To summarise, what (50) demonstrates is that: (i) Inalienable possession makes the expression of possessors obligatory; (ii) The *alatrenca* type of compounds excludes a dative clitic referring to the possessor; (iii) Dative clitics are allowed with verbs which are not compounds, and are obligatory for those verbs if IA is conveyed. From (i) and (ii) one can deduce that, when the *alatrenca* type of compound occurs, IA is expressed not by a clitic, but through another expression (i.e. a full NP).

Establishing that the theta-grid of *trenca* is [AG, TH], with a BENEFACTIVE as optional role, can account for the sentences in (50e-f) and (50i), the first two with a BENEFACTIVE, not the third. However, this cannot of itself account for the grammaticality judgements in e.g. (50j), where inalienability is crucial. Inalienability seems to be grammaticalised by establishing some kind of bond between nominals, the possessor and the inalienably possessed. I shall represent this relation by coindexing as in (51). This notation suggests an analogy between the phenomena just outlined and anaphoric binding which will be explored later.

- (51) *el caçador ala_itrenca [l'ocell]_i*
el caçador li_i trenca [l'ala]_i a [l'ocell]_i

For the moment, let us turn to some other sentences in which the *alatrenca* type appears.

- (52) a. *L'ocell es trenca l'ala.*
the bird ES breaks the wing
'the bird breaks his wing'
- b. *L'ocell s'alatrenca.*
the bird ES wing-breaks
(same meaning)

(52b) paraphrases (52a); *l'ocell* 'the bird' continues to represent the possessor of the wing, and *l'ala* 'the wing' the THEME, i.e. the broken entity. However, the context in which the compound is inserted is quite different from the one in (50) and (51). There is no AGENT for the action of breaking, here. The occurrence of the clitic *se/s'/es* correlates with the lack of an AGENT. Burzio (1986) relates Italian *si* to ergativity and obligatory reflexivity. These appear to be the two classes operative in

(52). A verb is said to be ergative when it lacks an external theta role and does not assign Case to its direct internal theta role (so that it must move to subject position to receive Case). The standard test for ergativity is the ability of the surface object to cliticise in *ne* in Italian, *en* in Catalan. Thus of the sentences above, (52b) is ergative while (52a) is not:

- (53) **Se'n trenquen l'ala molts.*
 SE EN breaks the wing many
 'Many of them break their wing'

Se n'alatrenquen molts.
 SE EN wing-break many
 'Many of them wing-break'

The *se* in (52b) corresponds then to an ergative *se*¹⁰. The *se* in (52a) seems to correspond to what Burzio calls "obligatory reflexives", which occur with "a small class of verbs like *sognarsi*, *immaginarsi* (...), which are transitive and obligatorily require the presence of a reflexive dative benefactive" (Burzio 1986:42).

What follows from this analysis is that (52b) can be generated as a projection of the verb *trençar* as considered up until now (i.e. as having a theta-grid [AG, TH]). For the generation of (52a) one has to appeal to another lexical entry, that of *trençar-se*, with theta-grid [EXP, TH]. This situation parallels the one in Italian, where *sognare* and *sognarsi* have different lexical entries, according to Burzio's analysis. Without postulating a lexical entry for *trençar-se*, it is not clear to me how to explain the lack of an AGENT theta role -- as we have seen, *es*, here, cannot absorb agentivity, since it stands for a benefactive. How to relate *trençar* and *trençar-se* in the lexicon is an aspect that I shall not investigate here¹¹.

It is worth noticing that the presence of a clitic linked to the possessor (*l'ocell* 'the bird') is obligatory in (52a), in the same way as in (50b). Also, the Case theoretic properties of *trençar* as Case assigner do not need to be altered for any of the analyses proposed: ergative *se* absorbs Case, thus (52b) is grammatical, and in (52a) the object gets Case from the verb.

There is another paraphrase of (52) which does not allow the presence of the compound:

¹⁰The only difference between ergative *se* and inherent reflexive *se* is that the first but not the second occurs with verbs which can have full NP objects (see Burzio 1986:42; compare (i) and (ii):

(i) *Il vetro si rompe.* 'The window breaks'
Giovanni rompe il vetro. 'Giovanni breaks the window'

(ii) *Giovanni si sbaglia.* 'Giovanni is wrong'
 **Giovanni sbaglia Pietro.*

¹¹Burzio's approach coincides in this respect with treatments of Italian clitic *si* which are syntactic instead of lexical for all other instances of *si* (cf. Manzini 1986).

- (54) a. A l'ocell se li trenca una ala.
 to the bird ES DAT breaks one wing
 'The bird breaks its wing'

A l'ocell se li trenquen les dues ales.
 to the bird ES DAT they-break the two wings
 'The bird breaks its wings'

- b. *A l'ocell se li alatrenca.
 (intended meaning 'a wing breaks to the bird')

(54a) is an instance of the so-called anticausative construction, in which the AGENT of the action expressed by the verb is not realised (or, rather, is absorbed by the clitic *se* in Romance in general). This can be seen by the oddness of the sentences in (55), where an adverbial implying agentiveness is included (this test can also found in Zubizarreta 1985).

- (55) ??A l'ocell se li trenca una ala voluntàriament.
 intentionally

The THEME complement is raised to subject position to get Case (since *se* is considered to absorb Case); the BENEFACTIVE remains in PP position and doubled with the clitic *li*, which as in all the sentences looked at until now is obligatory if inalienable possession is conveyed. Given that *una ala/les dues ales* is the surface subject of the sentence, and there is no surface object, the construction is ergative -- and *se* another instance of ergative *se*. This can be shown with *en* cliticisation of the subject:

- (56) A l'ocell, se n'hi ha trencat una, d'ala.
 to the bird SE of-them has broken one of wing
 'The bird has broken one wing'

While the description of (54a) is quite straightforward, a reason should be found to explain the ungrammaticality of (54b). (54b) is ungrammatical as an anticausative construction in which the THEME is part of the verb instead of being raised to subject position. Notice that there is no violation of the Projection Principle because all the theta roles in the verb's theta-grid are satisfied: the AGENT is absorbed in *se*, the THEME is realised in the compound, and the BENEFACTIVE is realised in the PP. However, there is no external theta role, and yet the clitic *se* gets (or absorbs) Case from *alatrencar*. This goes against Burzio's Generalisation, according to which "all and only the verbs that can assign a theta-role to the subject can assign (accusative) Case to an object" (Burzio 1986:178). Hence what is problematic here is the subject position, which is not assigned a theta role. In fact, the sentence becomes grammatical if interpreted as an example of *pro*-drop (meaning, for example, 'the body wing-breaks to the bird').

Finally, consider (57):

- (57) Això cor-trenca.
 this heart-breaks
 'This breaks one's heart'

The interpretation of this sentence is that there is a generic possessor of the heart; this null possessor is understood to be [+human] -- (57) cannot be interpreted as meaning 'this breaks any animal's heart', for example. In the sentence above, **El caçador alatrenca* 'the hunter wing-breaks', a [+human] null possessor is not possible, given our knowledge of the world, and the sentence is uninterpretable. To account for the meaning of (57), we can appeal to Rizzi's (1986) characterisation of arbitrary NP:

- (58) NP
 [+human]
 [+generic]
 [+/- PL]

In Rizzi's (1986) analysis, this can be realised in the syntax by *pro*, i.e. an empty category, which is in object position; the same seems to apply to (57). The value for the feature [+/-PL] is positive in Italian, and negative in Catalan.

The fact that the sentences above can be subsumed under the current theory without further specifications supports the idea that there is word-internal theta role assignment. The theta roles assigned in sentences with and without compounds remain the same, so it is possible to maintain that *ala* plays the same role in *El Joan li trenca l'ala a l'ocell* as in *El Joan alatrenca l'ocell*.

So far I have referred to the theta role saturated by *ala* 'wing' in the above examples as THEME. However, it is often pointed out that this label remains one of the most vague in theta theory. In the literature, some have taken THEME to correspond prototypically to entities which undergo movement (cf. Gruber 1976, Jackendoff 1976), or entities which change in state. Others consider THEME to stand for affected entity (cf. Anderson 1977). Rizzi (1986) distinguishes between "theme₁" and "theme₂" with respect to a semantic feature +/- affected. Themes₁ are affected. Themes₂ are unaffected, and are exemplified by the object of a process of perception or the theme of someone's psychological state. These two classes have been shown to be relevant for passivisation, middle formation and null object realisation, which are possible for theme₁, but not for theme₂, as illustrated in (59) to (61) (examples taken from Rizzi 1986:536ff).

- (59) a. the murder of John
 John's murder
 b. the sight of John
 *John's sight

- (60) a. John photographs easily.
 b. *John sees easily.
- (61) a. Gianni rende [pro felici]
 Gianni renders happy
 b. *Gianni vede [pro felici]
 Gianni sees happy

All the examples of the *alatrencar* type found in Catalan involve an affected THEME (i.e. a theme₁); *ala*, for instance, is a theme₁ with respect to *trencar*. On the other hand, newly coined examples with unaffected themes are of dubious grammaticality: ??*cama-veure* 'to leg-see', etc.¹²

Although Rizzi's point that the label "theme" identifies a spurious collection seems to be correct, I shall adopt the label THEME₁ only tentatively for my own purposes, since it is not clear that it is entirely effective for the semantics of Catalan compounds. An accurate conceptual definition of themes is a major topic.

3.3.2.1.1. On inalienability.

Two existing approaches to inalienability in syntax, which could be extended to Catalan compounding, are those of Jaeggli (1982) and Guéron (1984).

Jaeggli (1982) gives an analysis of this phenomenon for French and Spanish. His account is built on the assumption that inalienability involves the assignment of a special theta role, called θ_p , belonging to the theta-grid of the verb which "allow(s) the inalienable construction" (Jaeggli 1982:36). I have argued above, though, that inalienability holds between nominals. The analysis here will have more in common with that of Guéron (1984).

Guéron (1984) argues that inalienable possession constructions involve the formation of lexical chains, where "a chain is a set of two or more nominals related by anaphoric binding and interpreted as a single argument in LF (...). Unlike the chain which links a trace to its antecedent, a lexical chain contains more than one NP with lexical content. Under standard assumptions, lexical chains are thus formed by construal rather than by movement" (Guéron 1984:44). Lexical chains are subject to the following constraint (Guéron 1984:44):

¹²Likewise, only themes₁ seem to occur in the case of the compound type in section 3.3.2.2., *parallamps*; compare *talla-vidres* 'glass-cutter' with ??*veu-paisatges* 'landscape-seer'.

(62) Non-distinctness Constraint

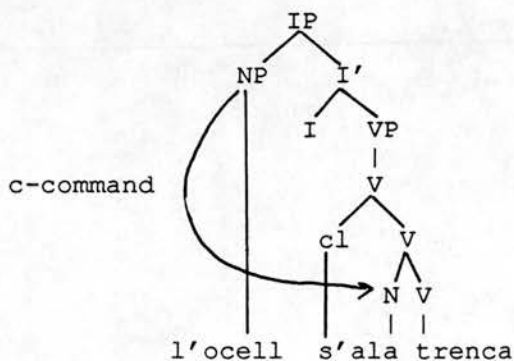
If A and B are links of a chain, then the referent of A is non-distinct from the referent of B.

(62) "permits [lexical chains] in which A denotes an individual and B is a feature compatible pronoun (...), or B denotes a body-part (...), but it rules out chains in which A denotes an individual and B denotes some object not intrinsically part of an individual" (op. cit.:45), i.e. the two cases allowed are a clitic and a coreferential NP forming a chain, or two nominals related by inalienability. Crucially, 'non-distinct' is *not* synonymous with 'identical', thus, 'John's arm' and 'John' are non-distinct, while clearly not identical. Identity and non-distinctness are different notions; those things which are identical are properly contained within those things which are non-distinct.

Guéron's claim is that links in a chain are related by anaphoric binding, indeed that "under the IA construal, a body-part NP is like a trace with respect to both binding and chain construction" (op. cit.:52); if so, there are some properties to be expected from IA constructions. Adopting this approach to compounding implies a modification of the assumption (cf. Sproat 1985) according to which only XP are liable to binding, since here I shall postulate coindexation of an X^0 category (e.g. *ala* in *alatrenca*) with another nominal category. The properties considered by Guéron appear in [1-5] and are illustrated with examples from Catalan in (63) to (67).

1. The possessor NP must c-command the N(P) designating the part of the body (or its trace).

(63)



Notice that in **El caçador trenca l'ala a l'ocell* 'the hunter breaks the wing to the bird', *l'ocell* does not c-command *l'ala* because of the intervening PP node. However, in *El caçador li trenca l'ala a l'ocell*, the dative clitic *li*, coreferential with *l'ocell*, c-commands *l'ala*, so that a chain is created between *l'ocell* and *l'ala* through the clitic. This might be an explanation for the obligatoriness of the clitic in this kind of construction -- and also an explanation for the lack of clitic in

sentences with *alatrenca*¹³.

2. The antecedent, i.e. the possessor NP, is obligatory. (This condition is not loosened for (57) above, since an empty object is postulated).

(64) *El caçador alatrenca.
the hunter wing-breaks

3. The antecedent must be in the minimal governing category¹⁴ of the N(P) denoting a part of the body.

(65) el nen_i veu com [la nena_j es cama_{*i/j}trenca]
the boy sees how_{IP} [the girl ES leg-break]
'the boy sees how the girl breaks her leg/s'

4. A fourth property, reported by Guéron after Kayne 1975, is that the inalienably possessed complement cannot be referential. They argue it then follows that this complement cannot be modified, in so far as modification implies referentiality. While this is certainly true of IA nouns in compounds, it does not seem to be the case otherwise in Catalan. Compare the French and Catalan examples:

(66) a. *L'ocell s'alatrenca la més blanca.
the bird wing-breaks the whitest (wing)

L'ocell es trenca l'ala més blanca.
the bird breaks its whitest wing
'The bird breaks its whitest wing'

b. *Je lui ai lavé les cheveux blonds.
I DAT have cleaned the hair fair
'I have cleaned his/her fair hair'

Je lui ai lavé les cheveux.
I DAT have cleaned the hair
'I have cleaned his/her hair'

The relation between referentiality and modification, therefore, seems to be more complex than suggested by Guéron. Truly, it must be assumed that not all links of a lexical chain can be referential if Principle C of the binding theory (that an R-expression is free in its minimal governing category) is to be satisfied. As for the lack of referentiality of compound-internal nominals, it is often presupposed

¹³For speakers who accept *El caçador trenca l'ala a l'ocell*, it may be that the preposition *a* is to be inserted for the purposes of Case-marking only, and that, in those circumstances, *l'ocell* still commands *ala*.

¹⁴Chomsky (1986b:169) defines a minimal governing category as follows: "A governing category is a 'complete functional complex' (CFC) in the sense that all grammatical functions compatible with its head are realised in it -- the complements necessarily by the projection principle, and the subject, which is optional unless required to license a predicate, by definition. (...) The local domain for an anaphor or pronominal α (..) is the least CFC containing a lexical governor of α -- the minimal governing category of α ".

that those elements are generic (i.e. non-referential), whether they designate inalienably possessed entities or not.

5. Because lexical chains are subject to the Theta Criterion, only one theta role may be assigned to the chain (this characterises lexical chains and sets a contrast with lexical anaphors -- compare (67a) and (67b)).

- (67) a. et_i menges una $cama_i$
 *BEN TH
 'you eat one of your legs'
- b. la $Maria_i$ es i renta
 AG TH
 'Mary washes herself'

What [5] implies is that, in my analysis of (50) above, *l'ocell* cannot be assigned a BENEFACTIVE theta role, because, if it is, a violation of the Theta Criterion results, because the lexical chain is assigned two theta roles, THEME and BENEFACTIVE. This analysis is consistent with the lexical entry of the verb *trencar*, which remains [AG, TH], with a BENEFACTIVE optionally realised. Compare (68a) and (68b).

- (68) a. El Pere li ha trencat la sabata al veí.
 DET Pere DAT has broken the shoe to the neighbour
 'Pere has broken the neighbour's shoe'
- El Pere ha trencat el vidre.
 DAT Pere has broken the window
- b. El caçador li ha alatrencat l'ocell a la nena.
 the hunter DAT has wing-broken the bird to the girl
 'the hunter has wing-broken the bird to the girl'

In fact, (68b) strongly supports the analysis in which *l'ocell* is not the BENEFACTIVE of *alatrencar*, because *a la nena* 'to the girl' cooccurs with it, and this last PP is to be interpreted as the BENEFACTIVE itself, like *al veí* is in (68a). In conclusion, *l'ocell* is identified as possessor (and technically not the benefactor) by coindexing, and a BENEFACTIVE theta role is allowed, but not required -- as optional theta role.

If the assumption is made (as will be argued later) that a verb which can assign Case must do so, a further prediction is made. Since *trencar* can assign Case, it must do so, and thus the presence of an NP is required. However, since the theta role THEME cannot be assigned to this NP, because it has already been assigned to *ala*, there is no theta role that this NP can be assigned¹⁵. This results in a violation of the Theta Criterion, by which an argument NP must bear a theta role. The only way for the sentence to respect the Theta Criterion is for *l'ocell* to get a

¹⁵The BENEFACTIVE is realised as PP by a Canonical Structural Realisation statement.

theta role. This is what happens when this NP forms a lexical chain with *ala* (i.e. *l'ocell* is theta-marked in virtue of belonging to a chain with a theta-marked link). Two nominals can only form a lexical chain if they can be interpreted as non-distinct. As a consequence, the fact that no alienably possessed nominals can appear in the *alatrenca* type of compound is predicted. In fact, the proposal here makes predictions which are notably more adequate than those making reference to inalienable parts of the *body*. Consider (69).

- (69) Aquest riu i el Danubi s'aiguabarregen a Alemanya.
 this river and the Danube SE water-mix in Germany
 'This river and the Danube join in Germany'

A grammar taking the noun in $[N V]_V$ to be an inalienable part of the body (as in (49)) would fail to characterise (69) as grammatical. However, a river and its water are non-distinct, and thus (69), in our account, is predicted to involve inalienability and to be grammatical.

The Non-Distinctness Constraint allows for lexical chains to be formed by nominals which have identical referents. Yet, (70) is ungrammatical (with the intended meaning):

- (70) *El caçador alatrenca l'ala.
 the hunter wing-breaks the wing
 'The hunter breaks the wing'

Therefore, some filter must stop this kind of lexical chain from being formed by identically referring nominals. This filter should not hold for the relations between clitics and nominals, if construed as chains, since e.g. the clitic doubling in (71) is good.

- (71) Li dic la veritat a la Maria.
 DAT I-say the truth to DET Maria
 'I tell Mary the truth'

Constructions analogous to (70) are also ungrammatical in languages such as Greenlandic, while grammatical in Yup'ik Eskimo (examples taken from Sadock 1986):

- (72) a. *276-inik ammassa-nik ammassa-ttorpoq.
 276 INSTR-PL sardine INSTR-PL sardine eat-IND-3s
 'He ate 276 sardines'
 (Greenlandic)

- (Yup'ik)

(73) If A and B are links of a chain, and they have lexical content, then the referent of A {can/cannot} be identical to the referent of B.

Overall, the predictions made by an analysis of inalienability in terms of the binding theory appear to be correct for the Catalan data here. The question that then arises is: Is this analysis compatible with the conception of words as anaphoric islands? If incompatible, how can other properties of words, so far accounted for on the grounds of anaphoric islandhood, be explained?

For a similar phenomenon, namely that of *self* prefixation (or compounding) in English, an analysis respecting the Lexicalist Hypothesis is developed in Di Sciullo and Williams 1987. For them, *self* binds the AGENT and THEME of the item it

attaches to, but not in the syntax; rather, the coindexing takes place in the theta-grid of the element that *self* is attached to, as in (74).

- (74) educated [AG, TH] / self-educated [AG_i, TH_i]
 denial [AG, TH] / self-denial [AG_i, TH_i]

So they argue that "this kind of argument binding [is] an operation on argument structures of lexical items" (op. cit.:59). Baker (1988b) remarks that this type of binding (binding of two arguments) is quite different from the one postulated by the authors earlier in the same book (binding of an argument by an affix); in particular, what the Di Sciullo and Williams treatment involves is the introduction of a syntactic process (binding) in the lexicon.

Independently of such issues of internal consistency, I shall argue that if the "syntactic" theory of binding can account for the binding of word-internal elements, suppressing this account unnecessarily diminishes the generality of the binding theory. That is, if the binding theory can cover all these cases (binding of lexical chains, of *self*, etc.), rather than reducing its predictive power because of the Lexicalist Hypothesis, it seems at least as interesting to renounce the Lexicalist Hypothesis itself. Along such lines, Lieber (1984) applies the binding theory in the analysis of *self*- words in English. Nevertheless, Lieber claims that rather than renouncing the Lexicalist Hypothesis altogether, it is possible to reformulate it by limiting its effect to some subcomponents (e.g. to the bounding theory but not the binding theory).

Taking the further step of dropping the Lexicalist Hypothesis altogether, the immediate issue arising is the source of ungrammaticality of paradigmatic cases of purported word anaphoric islandhood such as (75) (for similar examples, see Cabré and Rigau 1986:285, cf. Chapter Two).

- (75) a. *vaig comprar un obre-llaunes_i que e_i no s'obren bé
 I-bought an open-tins which no SE they-open easily
 'I bought a tin_i-opener which e_i don't open easily
 b. *es va cama_i-trençar la e_i que li feia mal
 SE leg-broke the (one) which DAT it-made pain
 'S/he leg_i-broke the one_i which hurt'

Notice that I consider these examples truly ungrammatical, and I cannot find any context in which they would improve. In this respect, then, I differ from Sproat and Ward (1987), who consider the following sentences of English and rule them out for pragmatic reasons only:

(76) Long-*legged* people don't like people with short *ones*.

Harry was looking for a *bookrack* but he only found racks for very short *ones*.

The sentences in (76) are, in the relevant respects, the same as (75), and illustrate typical anaphoric island effects. Because I find (75) ungrammatical and not unfelicitous, a pragmatic account such as the one in Sproat and Ward (1987) is unsatisfactory.

Compare (75) with (77), where no word-internal nominal is modified:

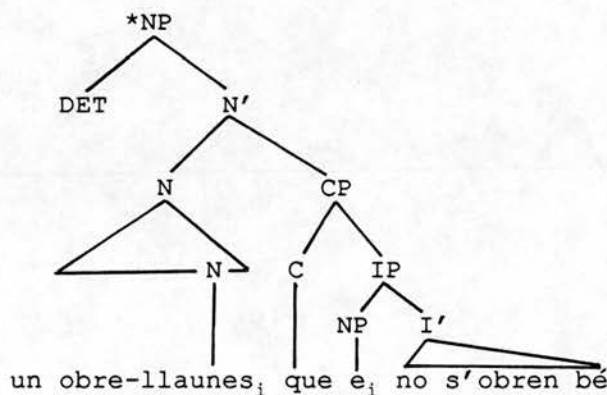
(77) a. Vaig comprar un obre-llaunes que no funciona.
I-bought a open-tins which no works
'I bought a tin-opener which doesn't work'

b. Vaig comprar un estri per obrir llaunes que no
s'obren gaire bé.
I-bought a tool for open-INF tins which no
ES they-open quite well
'I bought a tool to open tins that don't open
easily'

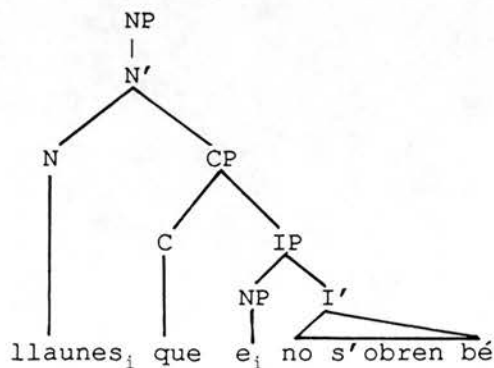
c. Es va trencar la cama que li feia mal.
'S/he-broke her/his leg which hurt'

The crucial difference between e.g. (75a) and (77b) can be located in the existence of an interpolated X^0 node in (75a) not appearing in (77b):

(78) a.

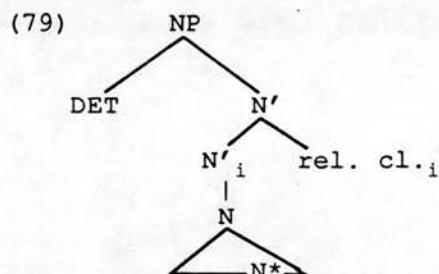


b.



What can account for the ungrammaticality of (78a)? The simplest explanation for this ungrammaticality is that the structure in which *llaunes* 'tins' is inserted does not allow for it to be modified. In effect, modification occurs between sisters, and *llaunes* is not a sister of the relative clause; rather, the whole compound is a sister of the relative clause, and therefore the only interpretation of, say, (77a) is one in which the relative clause modifies the compound, not one of its constituents.

Supposing we construed modification as a form of coindexing of modifier and modified constituents. Then, two sisters which were not related by theta-marking would be coindexed to be interpreted in a modification relation as in (79).



If a coindexing indicating modification was specified on the non-head daughter N^* , the structure would result in a violation of Chomsky's (1981:212) i-within-i condition (80a). More generally, a coindexation of N^* with the relative clause can be seen as conflicting with the A-over-A Condition, formulated by van Riemsdijk and Williams (1986:20) as in (80b):

- (80) a. i-within-i Condition
 $*[\gamma \dots \delta \dots]$
 where γ and δ have the same index.
- b. A-over-A Condition
 In a structure $_A[\dots _A[\dots] \dots]$, if a structural description refers to A ambiguously, then that structural description can only analyse the higher, more inclusive, node A.

A further prediction that we can make is that (81b) is going to be at least slightly worse than (81a) ((81a) is equivalent, in the relevant aspects, to (75a)).

- (81) a. **vaig comprar un para-brises_i [de mar]_i*
 I-bought a stop-breezes of sea
 'wind-screen'
- b. ***vaig comprar un neteja[para-brises_i] [de mar]_i*
 clean stop breezes
 'wind-screen-cleaner'

This is so because the noun *brises* 'breezes' and its modifier are even further apart than in (81a). (The ungrammaticality of the examples in (75) if generated by the application of Move α will be commented on later).

To summarise, it has been proposed that the binding theory applies inside words as well as outside them. (What remains true is that no NP pronouns and anaphors are part of compounds, so that the ungrammaticality of *ell-trencar* 'him-break', etc., is expected). Also, under the analysis here no reference needs to be made to the fact that the noun in [N V]_V compounds in Catalan is an inalienable part of an entity. Once it has been established that (i) Theta role assignment occurs inside compounds, and that (ii) Case assignment is obligatory if possible (I shall come back to this in Chapter Four), it follows that an object of the compound verb must appear, and that for it to get a theta role it must form a chain with the noun in the compound. The inalienable possession reading is then the only one available, given the parameter setting for the interpretation of lexical chains in (73).

3.3.2.2. The *parallamps* type

The *parallamps* 'stop-lightning', 'lightning conductor' type also presents internal theta role assignment; in particular, the noun in the construction corresponds to the THEME (THEME₁ in fact) in the verb's theta-grid¹⁶. This section will be devoted to the *parallamps* type and to other expressions which, together with it, have been grouped under the label "exocentric", such as:

(82)	un cent-peus a hundred feet	'centipede'
	un cap de trons a head of thunder	'scatterbrain'
	entre-cella between eyebrow	'space between the eyebrows'
	sota-barba under beard	'double chin'

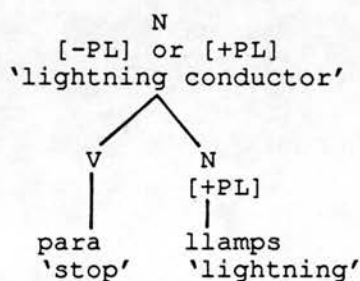
I shall not include among them compounds of the type of *Austria-Hongria*, treated in Chapter Six, for reasons which will become clear as I proceed.

The first noticeable characteristic of *parallamps* is that it does not designate a kind of lightning, nor the event of stopping something, but rather 'something which stops the lightning', i.e. a lightning conductor. The compound predicates of the lightning that it is stopped and, also, that there is something that stops it, thus two predications are being expressed. The meaning of *parallamps* correlates with the

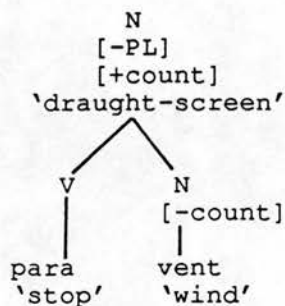
¹⁶Fabb (1984) and Roberts (1985) argue that INFL is necessary to license theta role assignment. This idea runs into problems in compounding of the kind of *parallamps*, where *llamps* is assigned a theta role without there being a compound-internal INFL node. Alternatively to Fabb's and Roberts' proposal, perhaps Case assignment -- instead of theta role assignment -- could be made dependent on the presence of INFL.

fact that, unlike *alatrencar*, neither of its constituents can be considered the head of the compound. That is, applying the IS A condition as a proof of headness (cf. Allen 1978), for instance, we have to conclude that *para* is not the head, nor is *llamps*, because *parallamps* is not an instance of stopping (of *parar*) nor is it one of lightning (of *llamp*). The fact that both *llamps* and *parallamps* are nouns is accidental, and this can be demonstrated by giving in some detail features other than the ones of syntactic category, as in (83).

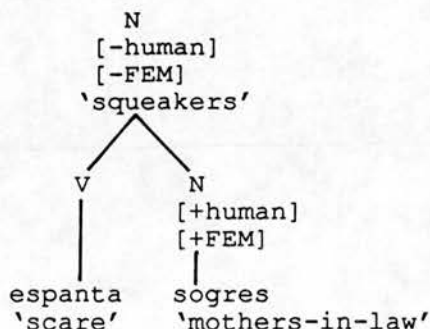
(83) a.



b.



c.



In (83a) the daughter noun is plural, but the compound can either be singular or plural; in (83b) the daughter noun is not countable, while the mother is countable and singular; in (83c) daughter and mother nodes have different values for the features gender and human. All these features are relevant for the determination of the well-formedness of the sentences in which the compounds appear and, therefore, must be specified in the representation of each compound. Taking the daughter noun as head does not allow us to predict the features of the mother, and thus must be dismissed. (This invalidates the analysis which Selkirk proposes for English compounds such as *cutthroat*, where *throat* is considered the head in the structural analysis of the compound -- see Selkirk 1982:25ff).

The following paragraphs discuss the properties of the *parallamps* type with respect to syntactic category, gender and other inherent morphosyntactic features, and internal structure.

Parallamps, *entrecella*, *sotabarba*, *cent-peus*, and the like are all nouns. The complex expression *cap de trons* can be used as noun and as adjective too (e.g. *Es un cap de trons* or *Es una persona cap de trons* 'He is scatty' and 's/he is a scatty person' respectively). I shall concentrate mainly on the noun type, assuming that adjectives can result by category change from nouns.

As noted, the morphosyntactic properties of the compounds are not determined by their daughter nouns. Quite often a compound of gender masculine contains a noun of gender feminine. I have not been able to find any feminine compounds of this type whose daughter noun is masculine. However, it is possible to find compounds of feminine gender with a feminine daughter. In fact, variation reported in the Fabra (1983) dictionary is allowed, as shown in (84); most interestingly, comparing this dictionary with Fabra's (1932) dictionary it is easy to find examples of compounds of this type whose value for the feature gender has changed.

- (84) porta-estandard [+FEM] or [-FEM]
 carry flag
 'flag holder'

As for the possible compounds *sotaorella* and *sotanàs*, both (85a) and (85b) seem possible to me, while (85c) is out.

- (85) a. el sota-orella
 the[-FEM] under-ear[+FEM]
 b. la sota-orella
 the[+FEM] under-ear
 c.*la sota-nàs
 the[+FEM] under-nose[-FEM]

It is always possible for a new compound to get masculine gender (except when it designates a female, in which case it is obligatorily marked as [+FEM]). Masculine is the unmarked gender in Catalan (e.g. *Homes i dones són humans* [-FEM]/**humanes* [+FEM]) and, therefore, it can be assigned by default in the absence of feature percolation in the formation of these compounds. Yet (85b), where the daughter noun is feminine, escapes this default specification.

With respect to the possibility of specifying the newly coined *sotaorella* 'under-ear' 'space under the ear' as [+FEM], the work of G. Corbett on gender (1990, forthcoming) is very suggestive. Corbett finds, across the languages of the world, two ways in which gender classes are organised: (i) a semantic one, where each gender class corresponds to a semantically based class (e.g. in Hindi); and (ii) a

phonologically based one, where each gender class has certain phonological characteristics. Corbett claims that French falls roughly within the second class, and I would say that Catalan does as well (for instance, most nouns ending in [ə] are [+FEM]). If this is so, we have an explanation of why *sotaorella* can be [+FEM]. Despite the fact that *orella* 'ear' is not the head of the word so that the percolation of its gender feature to the mother node is not possible, *orella* is the final element in the word, and hence the one that determines what the word "sounds like" for gender: it finishes in [ə], and therefore the word "sounds" [+FEM], and consequently can be classified as [+FEM].

This is certainly not an inviolable principle, but rather a tendency to which exceptions can be found (for example, *poeta* 'poet' ends in [ə] and it is [-FEM]). Indeed, the fact that we are dealing with a tendency accommodates perfectly well the data, with the possible variation in gender encountered (cf. (84) and (85)).

Turning now to number, like any noun, the compounds of this type are singular unless marked for plural (just like any noun). However, since many of them end in plural nouns (e.g. *llamps* in *parallamps*), the difference between singular and plural is phonetically neutralised. (The morphological behaviour of these compounds actually makes explicit the need for the determination of complex words; for example, masculine simple words ending in a sequence of vowel followed by /s/ regularly pluralise in -os, e.g. *país* 'country', *paisos* 'countries', but compounds ending in a plural form never get another plural marker).

In relation to internal structure the *parallamps* type has as constituents a verb (in the unmarked form of its paradigm) and a noun (either plural or singular), and the *sotabarba* type is constituted by a preposition and a noun (usually in singular). The nouns can be thought of as the complements of the verb and the preposition, respectively. However, these "complements" differ from complements in the X-bar theoretic sense in so far as they are X^0 categories, not XP. This is problematic for the analysis of exocentric compounds given in Williams 1981b and developed in Di Sciullo and Williams 1987.

In these, exocentric compounds are generated by "headless rules", which are unary rules:

(86) word \rightarrow phrase

Uses can be exemplified in (87) ((86) and (87) are taken from Williams 1981b:247, 250).

(87) a. N \rightarrow VP

[push_v up_p]_N

b. $V \rightarrow N$ (+phonological rule)

[breathe]_V; [breath]_N

Thus, exocentricity is equated with category change. The rule (88) is used in Di Sciullo and Williams 1987 for the French expressions in (89).

(88) $N \rightarrow XP$

- (89) a. V+N: essuie-glasse
wipe glass 'windscreen wiper'
- V+A: sent-bon
smell good 'tansy'
- V+Adv: couche-tard
lay down late 'late bedder'
- b. trompe-l'oeil
deceive the eye 'illusion'
- hors-la-loi
outside the law 'outlaw'
- arc-en-ciel
arch in sky 'rainbow'
- homme-de-paille
man of straw 'strawman'

Despite the correspondence between the compounds of form $[VN]_N$ in (89a) and the ones of Catalan in (90a), there is no compound in Catalan like the ones in (89b) and, even as nonce forms, those in (90b) are ill-formed.

- (90) a. porta-cigars
carry cigars 'cigar case'
- talla-papers
cut paper 'paper knife'
- afarta-pobres
fill poor-PL (variety of red beans)
- b. *enganya-l'ull
deceive the eye
- *apaga-la-set
satisfy the thirst

So rule (88) overgenerates for Catalan. The origin of this overgeneration is that a phrase is inserted in (88) and, yet, in Catalan all compounds are made of X^0 projections, X being a major category. Therefore, DET and XP must be excluded. Leaving aside the question of whether Di Sciullo and Williams' solution is adequate for French, or whether all the expressions in (89) are compounds (see Chapter Six), it is clear that (88) is too unconstrained for Catalan.

In the spirit of Di Sciullo and Williams' approach, (91) would generate the only exocentric compounds of Catalan:

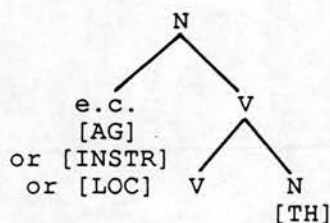
$$(91) N \rightarrow [[-N] [+N, -V]]$$

Here $[-N]$ and $[+N, -V]$ have to be read as bundles of features, the first designating both verbs and prepositions, the second nouns (see section 4.4.2.). What is not expressed in (91) is that the noun generated bears a thematic relation to the constituents of the compound, namely that it stands for the AGENT, INSTRUMENTAL or LOCATIVE of the verb. This semantic feature is the one which tells this type of compounding apart from category changes resulting in pairs such as *sal* 'salt' / *sal(ar)* 'to put salt on'.

It would be possible to associate the rewriting rule in (91) with an appropriate semantic operation, but the spirit of a constraint-based approach as GB is to define well-formedness conditions providing the basis of interpretation of all structures.

Suppose that the interpretation of the compounds at issue is taken as evidence for their underlying structure (and this is in fact a usual and legitimate procedure). Then we might want to specify this in the structural description of the compounds. Given that they always designate AGENTs, LOCATIVEs or INSTRUMENTALs, I shall postulate a noun head in their structure, even though this noun is not overt¹⁷. This empty category (e.c.) would be assigned a theta role by the verb, as in (92).

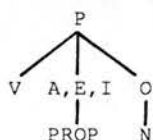
(92)



Notice that the theta roles which can be assigned to the e.c. are never internal theta roles in the verb's theta-grid: they correspond to an AGENT (or agentive), INSTRUMENTAL or LOCATIVE, i.e. to the external theta role in the verb's theta-grid, or to a semantic argument in Lieber's (1983) terminology. Thus consider the newly coined compound in (93).

¹⁷A precedent for an analysis of this kind is found in Bauer 1978, developed in the context of case grammar (cf. Anderson 1971); in this work the French exocentric compound *porteavions* 'aircraft carrier' and the English *cutthroat* are given the underlying structure in (i).

(i)



where A stands for agentive, E for experiencer, I for instrumental, and PROP for "pronoun+pragmatics".

- (93) roba-sabates
 steal shoes 'somebody who steals shoes' (AG)
 'something to steal shoes with' (INSTR)
 'place where shoes are stolen' (LOC)
 * 'somebody for whom shoes are stolen' (BEN)

The motivation for the structure in (92) assigned to compounds such as that in (93) is given by the fact that an external argument is realised outside its head's projection, and not as a sister of the internal arguments. Only semantic or external theta roles can be satisfied in the e.c. position. Given the assumption that word structures are binary, there is only one sister position of the verb in (92). If we agree that the internal arguments of a head must all hold the same structural relation with the head, it follows that *only* verbs with a single internal theta role to assign can appear in compounds, given binary branching. That this only argument satisfied in a compound must correspond to a THEME₁ has to be independently stated, and cannot be expected to be necessary (i.e. universal).

As the head, the e.c.'s features ([+/- animate], [+/- countable],...) percolate to the mother node. It is possible to ask what the nature of this e.c. is. First, it is base-generated, not the result of movement, and therefore it cannot be a trace. Second, it is not anaphoric, since it is free and does not pick up a preceding referent. As a consequence, it cannot correspond to PRO, which is anaphoric. It seems, then, that this e.c. is a pronominal and corresponds to a kind of *pro* of X⁰ level. If so, it obeys Principle B of the Binding theory ("a pronominal must be free in its governing category"; cf. Chomsky 1981:188) and must have its content identified locally by some element, be it agreement or something else. Suppose that SPEC-head agreement licensed this *pro*_N; then, in (94), *pro*_N would be licensed by its specifier, the compound-external determiner *un*.

- (94) *un*_i [*pro*_i [[*para*][*llamps*]]
 a stop lightning

This e.c. *pro*_N is licensed in much the same way that *pro* is licensed in object position (cf. Rizzi 1986), with a licensing principle like (95) (parallel to the one postulated in Rizzi 1986:519).

- (95) *pro*_N is governed by X_y⁰

where, in Catalan, X_y = DET. As in Rizzi's *pro* object, the features for gender and number of X are also those that are attributed to *pro*_N¹⁸.

¹⁸Catalan disallows the presence of *pro*_N without a modifier: **Vol un pro*_N 's/he wants a *pro*_N'. Not one of Rizzi's examples with a *pro* object lacks a modifier either. This is a topic for future research.

Other exocentric constructions can be postulated to have an e.c. pro_N :

- (96) un_i [pro_i [cap de trons]]
a head of thunder
'a scattered mind'
- els_i [pro_i [pobres]]
the-PL poor-PL

If (96) exemplifies this pro_N , it differs from (94) in that it is not theta-marked by a verbal sister; yet it is predicated of by its sister constituent. Alternatively, it could get a theta role assigned externally by the adjective *pobres* or similarly by *cap de trons*; that adjectives assign an external theta role is assumed in e.g. Levin and Rappaport 1986.

It is not clear to what extent such an e.c. should be put into use to account for some instances of category change. I have only found evidence for it to appear as head with a [V N] or [P N] complement. Despite the tentative character of the present formulation, a solution in terms of pro_N seems to be more promising than one along the lines of Williams 1981b, because it gives a unified treatment to all the constructions which have been termed exocentric and exemplified in (82).

I agree with Fanselow (1988:98) that empty elements have been used in word formation "in order to save some apparent generalization of feature determination from empirical counterexamples" and that "no theory of licensing conditions for empty affixes is in sight"¹⁹. I do not think that this applies to the empty category postulated here, namely pro_N , the motivation for which lies in theta theory: pro_N is postulated because the interpretation of the compound in (93), for instance, is accounted for if assignment of a theta role to an e.c. takes place.

An issue which I have not addressed yet is the order of elements in this type of compound. The endocentric compounds of Catalan are right-headed and this is encoded in a percolation convention (see section 4.4.2.). However, e.g. *parallamps* does not display right-headedness in any obvious fashion. In fact, it is argued by Williams 1981b and Di Sciullo and Williams 1987 that this type reflects the phrasal order of constituents, whereby, in French, as in Catalan, complements usually follow verbs, and prepositions are followed by their complements, as in (97).

- (97) Zeus va parar el llamp.
Zeus stopped the lightning
- No miris sota l'escala.
Don't look under the stairs

¹⁹Gavarró (forthcoming) argues against some particular instances of zero affixation posited to save the Righthand Head Rule of Williams 1981b.

Having postulated an empty head, I do not find any empirical evidence to decide whether it is word-initial or word-final. On the other hand, it is evident that the two elements which surface, e.g. *para* and *llamps*, or *sota* and *escala* (in *sotaescala* (space under the stairs)) are an instance of left-headedness (*para* and *sota* are the heads, and the nouns their complements). On these grounds, my suggestion is that "exocentric" compounds are consistently left-headed.

Leaving aside the historical evidence according to which these compounds originate in an imperative clause in Latin, with the head verb on the left, a functional reason for exocentric compounds to be left-headed comes to mind. Given that the rest of compounds of Catalan are right-headed, keeping a different class apart by the order of constituents (i.e. by an overt characteristic) is a way of having its internal structure made transparent. (The need for this transparency is what motivates the Principle of PF Interpretation of Baker (1988a), discussed in Chapter Four).

Chapter Four

Case Theory

4.1. A generalised theory of Case

The previous chapter has argued for the relevance of theta theory in compounding. Given the Visibility Condition in (1), it then seems that word-internal Case assignment is essential to the full characterisation of well-formed expressions.

- (1) The Visibility Condition
An element is visible for θ -marking only if it is
assigned Case.
(Chomsky 1986a)

The "identificatory" function of Case has a long tradition in linguistics, which is briefly sketched in Fillmore 1968. In Baker's (1988a) formulation, Case (g-Case, or generalised Case, hereafter) is a generalisation over a series of different overt grammatical properties, all of them having as goal the identification of argument and modifier relationships. The overt relationships which Baker refers to as manifestations of g-Case include verbal agreement, morphological case, directional adjacency and incorporation (in the technical sense of Baker 1988a). These four instantiations of g-Case are exemplified in (2), (3), (4) and (5) respectively:

- (2) Wi:rv:n wa- hra- kv- ? tsi-r.
William AOR 3MASC-SG 3NEUT-O see dog
'William saw the dog'

Tsi-r wi:rv:n wa-hra-kv-?.

Wa-hra-kv-? wi:rv:n tsi-r.
(same meaning)

(Tuscarora)¹

¹This example is taken from Baker (1988a), who quotes it from Williams' (1976) thesis on this Iroquoian language.

- (3) Graecia capta ferum victorem capit.
 GreeceNOM conqueredNOM fierceACC winnerACC captivates
 'Conquered Greece captivates the fierce conqueror'

Ferum victorem Graecia capta capit.

(Latin)

- (4) William saw the dog.

*William saw unexpectedly the dog.

- (5) I?i k- atat-nuhs- ohras.
 I 1SG-S REFL house wash
 'I washed my own house'

(Mohawk)²

(2) exemplifies a head-marking language (cf. Nichols 1986), where arguments themselves are not marked, but the inflected verb has agreement markers which tell us that *Wt:ru:n* 'William' (the 3MASC noun phrase) is the subject agent and that *tst-r* 'the dog' (the 3NEUT noun phrase) is the object patient. The two arguments are identifiable when they appear in varying positions in the sentence. In Latin, the order of phrases is also quite free, as exemplified in (3), but here it is the noun phrases which have morphological Case markers, and so are identified as subject and object. In the English example (4) directional Case assignment under adjacency is shown: the object patient of the verb has to appear immediately to its right; when an adverbial appears in between the two, the sentence is ruled out. Finally, in the Mohawk example in (5), the argument *nuhs* 'house' is identified in virtue of its incorporation to the verb.

Many languages have mixed systems of g-Case assignment, that is to say some cases are expressed by one means, and some by others. Baker gives as an example Turkish, in which the subject triggers verbal agreement, but has no morphological case marker, while the object -- if definite -- is morphologically marked as accusative, but there is no object/verb agreement.

Also, there are more complex constructions where g-Case is expressed by more than one device. For example, English pronouns, being somewhat exceptional in the context of the language itself, exhibit morphological case markers (e.g. *he* and *she* vs. *him* and *her*), but also their position with respect to verbs and prepositions (i.e. with respect to Case-assigning categories) is determined, with Case being assigned by verbs and prepositions to the right, and under adjacency.

- (6) I saw him.
 1NOM 3ACC

*Him saw I.

²Baker 1988a takes this example from Postal (1962:332).

*Me saw he.
'I saw him'

The same can be found in Catalan clitics (and Romance clitics in general), this being one of the few areas where morphological case persists.

- (7) Jo miro.
1NOM I-look
'I look'
- Em miro.
REFL I-look
'I look at myself'
- Em mires.
1ACC you-look
'You look at me'

In addition to morphological case, there is a manifestation of g-Case in subject/verb agreement (as exemplified in (7)). In consequence, some constructions present a certain amount of redundancy: in *he sees us*, for example, we have three clues on which to decide who does the seeing and who is seen: morphological case marking, subject/verb agreement, and word order (the experiencer precedes the verb and the theme follows it)³.

G-Case, as a grammaticalisation of semantic relations, must be linked to theta theory. Typically, the same categories which are theta role assigners, namely verbs and prepositions, are Case assigners. Categories which are typically assigned theta roles, nominals, are Case-marked (though in head-marking languages the verb bears the marking, instead of its arguments). However, the particular relation between g-Case assignment and theta role assignment varies from language to language. In this connection, Baker offers the following typology:

- (8) a. Semantic Case
Case marker correlates with a single theta role
- b. Inherent Case
Case marker correlates with some subset of theta roles
- c. Structural Case
Case marker uncorrelated with theta roles

This is an idealised typology, and so it has to be understood that existing languages may display semantic Case in many of their constructions, though not all of them, and so on. One finds languages like Estonian (cf. Baker 1988a:113) in which there is a biunique relationship between each particular theta role and each Case; this is called semantic Case (because the semantics -- the thematic relation -- is

³Of course, there is no reason to presuppose that natural languages tend to avoid redundancy; in fact, it has been often suggested in the literature that they favour it (e.g. Sadock 1983).

unequivocally inferrable from Case). This relationship can be loosened to give inherent Case, where Case indexing and theta role assignment still coincide, but where one Case possibly materialises more than one theta role. Finally, all theta role assigned constituents are Case-assigned, but not necessarily by the same head.

This minimally respects the Visibility Condition, (1) above, revised by Baker as in (9):

- (9) The Visibility Condition (revised)
B receives a theta role only if it is Case-indexed.
(Baker 1988a:117)

Baker's modification, in its use of the expression "Case-indexing", is intended to strengthen the idea that g-Case as an abstract relation holds at S-structure and is only spelled out in PF as morphological case, agreement or whatever. G-Case indexing is meant to be universal, while its manifestation is parameterised.

This conception of g-Case is expressed in the Principle of PF Interpretation, (10), which guarantees that no (abstract) Case will go phonetically uninstantiated. This does not require, however, that all abstract Case distinctions (between, for instance, accusative and oblique Case) be kept on the surface, which is required in other approaches to Case (cf. Bach 1983).

- (10) The Principle of PF Interpretation
Every Case indexing relationship at S-structure must be interpreted by the rules of PF.
(Baker 1988a:116)

Presumably for functional reasons, natural languages seem to follow the generalisation (11) below, which ensures that the argument identification is generally unambiguous:

- (11) A single item cannot Case-index two NPs in the same way.
(Baker 1988a:122)

A few exceptions to this generalisation have been found in natural languages, but they are certainly rare. In those systems, no formal distinction is made to signal arguments bearing different theta roles.

In the context of word formation, Baker (1988a:122) assumes another principle:

- (12) The Case Frame Preservation Principle (CFPP)
A complex X of category A in a given language can have at most the maximal Case assigning properties allowed to a morphologically simple item of category A in that language.

Still, it is possible to try to derive the CFPP from other principles of the grammar, including percolation principles in word structure (see section 4.4.2.).

4.2. Compounding and Case theory

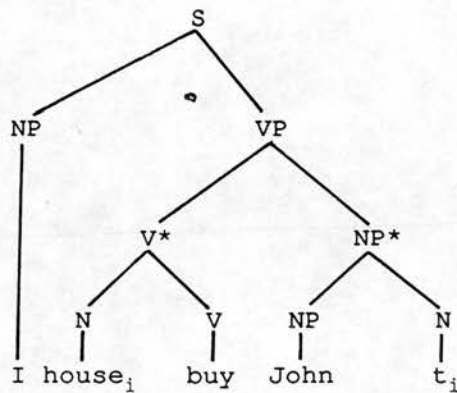
Baker's reconstrual of Case accommodates contexts for which the Case Filter is no longer applicable⁴.

- (13) Case Filter
 *NP if NP has phonetic content and no Case.
 (Chomsky 1981:49)

So, in his analysis of noun incorporation as X^0 movement, the noun which incorporates does not have to receive Case to satisfy the Case Filter (because incorporation is, in itself, an instance of g-Case) and the Case which the verb usually assigns can be assigned to another element, as in (14) (example from Baker 1988a:118):

- (14) a. Wa- hi- nuhs- ahni:nu: John.
 AOR 1SG-S-3MASC house buy John
 'I bought John's house' (Oneida)

b.



Here *wa-hi-nuhs-ahni:nu:* 'housebuy' assigns Case to *John* (in fact the Case that 'buy' assigns to 'house' in 'I buy a house'), because *nuhs* 'house' is phonologically identifiable at PF after incorporation, without Case being assigned to it by the verb; in this way, there is Case left to be assigned to the possessor. If there is noun incorporation, conventional Case assignment to the incorporated noun is superfluous.

The situation in Oneida resembles that of Catalan compounds:

⁴As noted, satisfaction of the Case Filter entails satisfaction of the Visibility Condition. But, given Baker's approach, respecting the Visibility Condition does not imply respecting the Case Filter. The Case Filter refers to a particular instance of Case-indexing, Case-indexing of NPs in a sentential context.

- (15) a. El Joan li trenca la cama a la nena.
 DET Joan DAT breaks the leg to the girl
 'Joan breaks the girl's leg'
- b. El Joan camatrenca la nena.
 DET Joan leg breaks the girl
 'Joan breaks the girl's leg/s'
- c. *El Joan camatrenca a la nena.
 DET Joan leg-breaks to the girl

In (15a) *la cama* is assigned Case by *trenca* 'breaks' and the possessor *la nena* 'the girl' is assigned Case by the preposition *a* 'to'. *La nena* in (15b) must be assigned Case because otherwise the sentence would contain a violation of the Visibility Condition; the only constituent which can possibly assign Case to it is the compound *camatrenca*, which has presumably inherited the Case-assigning capacity of *trenca*. I argue later that the ungrammaticality of (15c) arises through the failure of the verb to assign Case, which must be assigned.

Mithun (1984), in her study of noun incorporation (in the traditional sense, not the technical sense of Baker) considers four types of noun incorporation (NI hereafter).

- Type I: In lexical compounding, "a V and a N combine to form a new V"; the noun bears a particular thematic role with respect to the verb, and the compound is interpreted as referring to a common, socially relevant activity; the noun is generic as opposed to definite.
- Type II. This NI has as effect an extra change in sentence structure: the resulting compound verb retains its Case-marking capacity (as in the Oneida example above).
- Type III. This NI results in a manipulation of discourse, since its main characteristic is that nouns are made into background information by being incorporated, leaving more salient positions (object positions) for focused elements.
- Type IV: In classificatory NI, the incorporated noun acts as a generic modifier of the verb, leaving room for fine-grained semantic classifications of the event denoted by the simple verb.

The Catalan examples seem to correspond to type II, the type in which "when a transitive verb incorporates its object, then an instrument, location or possessor may assume the vacated object role" (Mithun 1984:856). Mithun claims that there is a hierarchical implication between the four types of incorporation, so that a language with type IV incorporation will also display types III, II and I, a language

with type III will also have types II and I, and so on. However Catalan does not present any compounds of type I, at least not with the structure $[N V]_V$. Baker (1988a:23, note 4) suggests that type II should be conflated, depending on the language, either with type III or with type I. I shall favour for Catalan the latter of Baker's positions, instead of entirely dismissing the hierarchy established by Mithun.

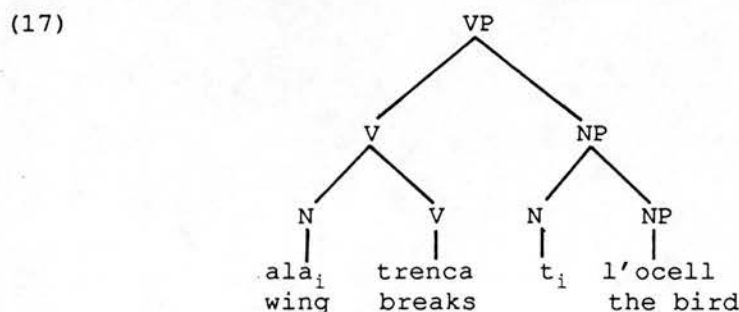
Given the striking parallelism between the Oneida example and the Catalan one, is it feasible to analyse the Catalan compound as noun incorporation, in particular in the manner proposed by Baker (1988a): N to V movement in the syntax? What evidence can be adduced for and against this hypothesis?

4.2.1. Why not Move α ?

In Chapter Three it was established that the noun which appears in $[N V]_V$ compounds in Catalan corresponds to the THEME in the verb's theta-grid. This THEME originates as direct object at D-structure and can surface as direct object (i.e. as the argument which is Case-marked by the verb) in syntactic structures, but need not do so, as shown in (16).

- (16) a. El Joan li trenca el cor.
 DET Joan DAT breaks the heart
 'Joan breaks his/her heart'
- b. El cor se li trenca.
 the heart SE DAT breaks
 'his/her heart breaks'

If the noun ordinarily surfaced as a direct object in all cases, it would be natural to provide a syntactic treatment in terms of Move α . Then, $[N V]_V$ compounds would be the result of incorporation in Baker's technical sense:



However, this is not the case. It is possible to generate compounds of this type without mentioning grammatical relations, by appealing only the information contained in lexical entries (on theta roles, Case, etc.). To generate, e.g., *cortrencar*,

and *alatrençar* it would be sufficient to state that the THEME direct object adjoins to the verb to form a compound verb. Furthermore, the reference to the theta role THEME₁ is not dispensable in favour of a formulation in terms of direct objects only. There are D-structure direct objects which cannot adjoin to a verb to form a compound; this is the case for the objects of unaccusative (or ergative) verbs (cf. Burzio 1986). For example, *arribar* 'to arrive' has an internal argument, which is the verb's direct object at D-structure, and becomes its subject at S-structure to be assigned Case. The compound in (18) is ungrammatical, and it presents a noun that would otherwise be the (D-structure) direct object of the verb.

- (18) *noi-arribar
 boy arrive

(18) is ruled out by making reference to theta roles: the noun in the compound is assigned a theta role different from THEME; *noi* 'boy' corresponds to the AGENT, as shown by (19), where an agentive adverbial is included (adverbs like *voluntàriament* can only occur with agentive verbs).

- (19) Els nois arriben tard perquè volen.
 the boys arrive late because they-want-to
 'the boys arrive late voluntarily'

In conclusion, to adequately generate the [N V]_V compounds of Catalan, reference to theta roles is both necessary and sufficient. Reference to direct objects is in itself insufficient, though it cannot be excluded straight away (if it is taken to account for the data together with theta theoretic notions).

What arguments can be used, then, to choose between an analysis with Move α and one without? First, I shall look back at the evidence that Baker (1988a) uses to postulate the application of movement for the NI that he treats in this way. There are four arguments which I shall examine in turn.

First, NI is referentially transparent, i.e. the noun can be interpreted as referential and definite. This transparency is attributed to the presence of a trace (the noun trace) outside the word. But Catalan compounds are not referentially transparent. The fact that the noun in (20a) can be considered definite, is an effect of the speaker's knowledge of the world, not a characteristic of the compound itself, as (20b) shows.

- (20) a. La Maria es colltorça.
 DET Maria SE neck-twists
 'Maria's neck twists'
 b. El Joan alatrenca la mosca.
 DET Joan wing-breaks the fly
 'Joan breaks the fly's wing/s'

In (20a) *coll* must necessarily refer to Maria's neck. However, this is the only interpretation available because we know that people have only one neck. That the compound's noun is not definite in itself is proved by the undeterminacy which arises in (20b): because flies have more than one wing, the sentence can be understood as meaning that Joan breaks more than one of them. On this occasion, our knowledge of the world does not tell us what specific event is being referred to.

Second, NI means incorporation of the head noun out of a noun phrase. Thus, there is the possibility of stranding of demonstratives, as in (21) (taken from Baker 1988b:12).

- (21) i?i k- nuhs- nuhwe?-s [thikv t]_{NP}
 I 1-SG-S house like that
 'I like that house'
 (Mohawk)

In Catalan, this gives rise to ill-formed expressions, such as (22a). In (22b) *aquella* can only be understood as the possessor ('that female'), not as a demonstrative specifying 'leg'.

- (22) a. *el Joan li cama_itrenca aquella t_i
 DET Joan DAT legbreaks that-FEM
 'Joan breaks that leg to her'
 b. El Joan camatrenca aquella.
 'Joan leg-breaks that one'

Third, if the noun in the Catalan compounds corresponds to the (THEME) object and only to it, incorporation of adjuncts and objects of prepositions is excluded -- because none of these are governed by the verb, i.e. none of these are true objects and, therefore, the trace the noun would leave after movement would be ungoverned, resulting in a violation of the ECP. Let us see if the predictions of the movement analysis are borne out. There is no possible test for the diagnosis concerning adjuncts, because there are no adjuncts which correspond to the THEME in a verb's theta-grid (i.e., adjunct implies not belonging to the theta-grid of the verb); so, for this matter, a movement analysis is not required, because it does not make any prediction different from an analysis in terms of theta theory. For the second diagnosis, we would have to find a Catalan verb with the following characteristics: (i) having a THEME in its theta-grid corresponding to an inalienable part of an entity; and (ii) not being able to assign Case, so that a preposition is required to assign Case to its complement. Then, Baker's prediction would be that that complement would not incorporate, while an analysis without the application of Move α would predict that the compound was grammatical. But, because THEMES introduced by prepositions are so rare themselves, I have been unable to find a way of establishing which prediction is fulfilled.

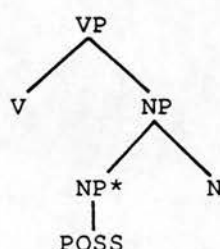
Likewise, another consequence of movement observed by Baker, namely that the noun which appears in object position as a result of dative shift cannot incorporate, cannot be used in Catalan, since there is no dative shift construction.

Given the evidence above, there seem to be no reasons to postulate a movement analysis; i.e. there is no phenomenon which could be accounted for in terms of movement and not otherwise. In actual fact, an account in terms of theta theory seems more economical, since no reference to grammatical relations is then made. A treatment without movement avoids the following complication in the assignment of Case in NI (Baker 1988a:111):

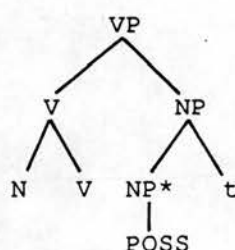
"Given that the verb governs the possessor NP* it is free to assign Case to NP* only if it does not need to assign its Case to the whole object NP. Since it does in fact Case-mark NP*, we conclude that NP does not need Case."

That is, in the configurations in (23) and (24), Case assignment follows different paths:

(23)



(24)



In (23), the head of NP gets the Case which the verb assigns; in (24), it is NP* which gets it, because according to Baker *t* does not need to be assigned Case. One of the shortcomings of this analysis is that it is not clear how to stop NP* from getting Case from the verb in (23); one needs to stipulate that NP* is not Case-assigned by the verb unless the head of NP does not need to be assigned Case. Problems such as this are, of course, avoided if the analysis chosen does not involve Move α . This is not an argument against incorporation in general, but it is an issue which further research on incorporation will have to tackle.

Since there is no necessity for an incorporation analysis of Catalan compounds, are there any arguments to exclude it? Over and above economy, some reasons to prefer an analysis without movement are presented in [1] to [4].

1. The existing and possible right-headed compounds of Catalan are of the structure [N N], [A N], [N A], [A A], [A V], excluding [N V]. None of these (with the possible exception of [A V]) would be generated via movement. If [N V]_V was generated by Move α , we would be left with two questions: (i) Why does Move α apply only to nouns incorporating into verbs, i.e. why do we not find verb incorporation or preposition incorporation, as in other languages? (ii) Why is [N V]_V not possible as a compound formed without Move α , given that verbs pattern like nouns and adjectives in their combinatory ability with all other categories? In other words, we would have to explain the gap in the list of Catalan compounds formed without Move α . If [N V]_V is formed without movement, then Catalan compounds are all of the form:

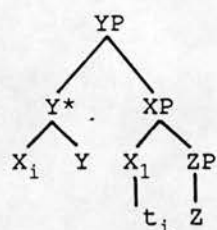
$$(25) \quad [Y X]_X$$

where $Y \in \{N, A\}$
and
X ranges over all major lexical categories

and [N V]_V is not exceptional in any respect.

2. Following Mithun's (1984) implicational typology of noun incorporation, types III and IV should only be possible in a language if types I and II are already displayed by the language in question. According to Baker, incorporation in the technical sense (or "full syntactic noun incorporation, i.e. movement of noun roots" Baker 1988b:23) is incorporation of types III and IV in Mithun's terminology. Then, if Catalan compounds were instances of such NI, and therefore of type III or IV, Catalan would be a counterexample to Mithun's implicational typology -- since [N V]_V compounds are the only candidates for NI in Catalan. So, an NI movement analysis of these compounds is inconsistent with Mithun's typological hierarchy.
3. It has been suggested in Chomsky 1988 that the application of Move α should be minimised in order to make derivations maximally economic. If the Catalan compounds at issue can be generated without the application of Move α , and by using devices required to generate other compounds anyway, the resource to movement seems superfluous.
4. My claim that Catalan compounding does not involve movement follows from the present formulation of the theory. That is, the structure in (26), typical of incorporation, constitutes a violation of the ECP, because the trace t is not properly governed. (Baker's notion of incorporation, which obviously rests on the well-formedness of (26), requires modifications of the theory that are not adopted here).

(26)



Since there is no evidence necessitating a movement analysis, I shall not resort to movement. In consequence, the claim is that Catalan does not display (noun) incorporation, in Baker's sense.

4.3. Catalan compounds and Case

Overall, what comes out of the approach to Case presented above is that word formation in itself is a means of expression of g-Case. Constituents are g-Case-indexed in virtue of belonging to a complex word, and according to Baker they are not subject to other Case-indexing mechanisms. For instance, the Case assigning capacity of *camatrençar* is identical to that of *trençar*; in the examples above both *cama* and its possessor are identifiable because the first is part of a complex item and the second is the surface object of the verb; at a general level then, both are g-Case-indexed.

Baker proposes a movement analysis of complex words with g-Case-indexing at S-structure; Case-indexing should certainly take place before PF (since this level of representation overtly expresses g-Case indexings). It is also understandable that certain kinds of g-Case indexing, such as that between subject and verb resulting in agreement, should be at S-structure, given that an NP may only be in subject position as a result of movement and, yet, has to agree with the verb. I also propose that g-Case indexing can take place in the process of word formation, particularly compounding in the syntax, but without movement. This g-Case indexing affects only a (nominal) X^0 category, if, as generally assumed, compounding involves only X^0 categories.

Taking into consideration the semantic restrictions imposed on Catalan compounds of the *camatrençar* type discussed in Chapter Three, the g-Case indexing at issue can be typologised as an instance of semantic Case; i.e. in Catalan compounds of this type the g-Case indexing constituent and the theta role assigner correspond to the same element, and the theta role assigned is always a THEME_1 .

From the distinction among various kinds of g-Case indexings, a principled classification of complex expressions emerges. Among the expressions which have been considered compounds (pretheoretically), some seem to manifest g-Case indexing in a way which is particular to words (that is, g-Case as compounding), others manifest g-Case indexings through standard phrasal Case markers. (27) and (28), respectively, seem to belong to these two classes:

- (27) a. cama-trencar
leg break-INF 'to break somebody's leg/s'
b. pasta dents
paste teeth 'tooth paste'
- (28) a. cuc de seda
worm of silk 'silkworm'
b. màquina de rentar
machine of wash-INF 'washing machine'

In (28), the preposition *de*, 'of', as in normal phrasal contexts, assigns Case to the noun that follows. In the syntax, the NP *poemes* in (29) is not assigned Case, and the whole sentence is ruled out by the Case Filter.

- (29) *Llegeixo un llibre poemes.
I-read a book poems
'I read a poetry book'

But *llibre poemes* has precisely the same form as *pasta dents*, [N N], which is well-formed, and does not, then, violate the Visibility Condition. Hagit Borer (1984) finds constructions such as (30), which she calls "wordphrases", and which are in all respects like those in (28) and contrast with those in (27).

- (30) [[take][advantage][of]]_v (NP)
coat of arms
pain au chocolat
bread to-DET chocolate (French)

She considers that they are "lexicalized items, with specific meaning" but that "syntactic conditions must be met inside them" (Borer 1984:27) and Case assignment by the dummy Case markers *of* and the French *à* are thus explained.

Similarly, for complex predicates like those in (31), Gràcia (1986) already indicates that Case assignment (in the narrow sense, i.e. satisfaction of the Case Filter) takes place normally. She relates this to the ability of the noun to pronominalise (since, in the analysis of clitics that she assumes, they absorb Case):

- (31) a. En Lluís feia petons a la nena.
 DET Lluís made kisses to the girl
 'Lluís kissed the girl'
- b. *En Lluís feia petons la nena.
 DET Lluís made kisses the girl

Her analysis explains why (31b) is ungrammatical: *la nena* has not been assigned Case. If *petons* did not need to be assigned Case, the Case assigning properties of *fer* 'to make' could be used to assign Case to *la nena*, and (31b) would be grammatical.

Despite the parallelism between the noun *cuc de seda* and the complex predicate *fer petons* (i.e. in both constructions Case is assigned as in phrases), there is a difference to point out. Gràcia attributes the possibility of pronominalisation to Case, but *seda* cannot pronominalise, unlike *petons* in (31), which can pronominalise as shown in (32):

- (32) En Lluís en feia a la nena.
 DET Lluís EN made to the girl
 'Lluís (kissed) the girl'
- En Lluís n'hi feia.
 DET Lluís EN DAT made
 'Lluís (kissed) him/her'
- (33) a. Vull un cuc de seda.
 I-want a worm of silk
 'I want a silk-worm'
- b. Vull un vestit de seda.
 I-want a dress of silk
 'I want a silk dress'
- c. En vull un de seda
 EN I-want one of silk
 'I want a silk one'
 * as pronominalisation of (a)

If, as I have maintained so far, Case assignment takes place in the same way in (33a) and (33b), we will have to come to the conclusion that such Case is not a sufficient condition to pronominalise. Then *de seda* does not alternate with a pronoun for reasons other than Case theory. Pronominalisation may be sensitive to lexicalisation. Evidence to this effect can be adduced from the difference between the complex predicate in (34) and those in (31)-(32):

- (34) a. Et passes el dia somiant truites.
 REFL you-spend the day dreaming trouts
 'You spend the day fantasising'
- b. Et passes el dia somiant-ne.
 REFL you-spend the day dreaming-NE
 'You spend the day dreaming of them'
 * as pronominalisation of (a)

Here, cliticisation of *truites* is not possible, because *somiar truites* is a *lexicalised* complex predicate.

Behaviour with respect to Case is a diagnostic for compoundhood. In some cases, however, no Case markers would appear in the corresponding syntactic structure, and so Case theory is of no use to tell apart compounds from "syntactic words" (in Borer's terminology). This is what happens with (35), which presents no internal Case markers and which once inserted in a syntactic structure behaves as any N' would behave. This construction is not sufficiently characterised by the subtheory of Case alone and will have to be assigned structure on the grounds of another subtheory.

- (35) *curt circuit*
short circuit

To summarise, it has been shown that, in Catalan, Case is assigned under adjacency to the right for ACC Case (as shown in (36))⁵, but that there is also g-Case assignment resulting in compounding, which does not fall within the scope of the Case Filter.

- (36) *Veu el Miquel.*
s/he-sees DET Miquel
'S/he sees Miquel'

**El Miquel veu.*
DET Miquel s/he-sees
(same meaning)

Compounding is realised internal to X^0 structures, while g-Case indexing among phrases is realised both in word order and subject/verb agreement.

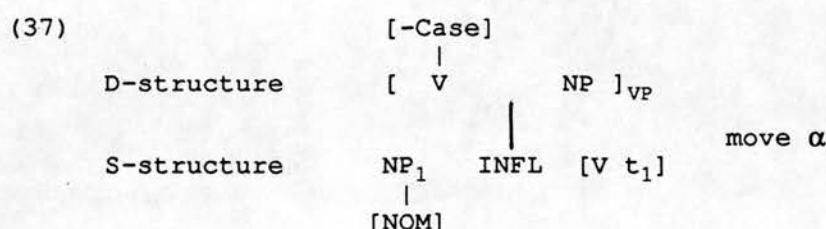
4.4. Case theory and X-bar theory

In connection with the topic of word order, I shall briefly discuss the interaction of Case theory and X-bar theory. They are related in that both of them can make reference to linear order. Case may be assigned, in languages such as English and Catalan, under adjacency and directionally. X-bar theory, for its part, sets limits on the occurrence of specifiers and modifiers and their heads; the parameterisation of the X-bar principles basically involves the determination of the order in which heads, modifiers and specifiers appear with respect to each other.

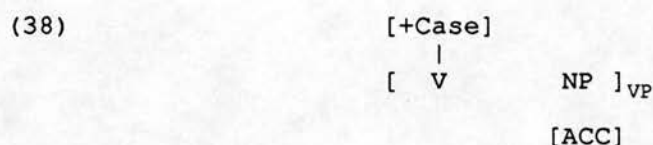
There is a certain amount of redundancy between the two theories. For instance,

⁵Adjacency needs to be loosened to allow for interpolated adverbials, which, unlike in English, are good in Catalan; e.g. *Veu cada dia el Miquel* 's/he sees everyday Miquel' 's/he sees Miquel everyday'.

given a language like Catalan, the direct object canonically follows the verb. Case theory states that a verb assigns Case to the right, if it does at all. Supposing that we have a verb which has no Case to assign (an unaccusative verb like *arribar* 'to arrive', for example); the NP which follows the verb at D-structure has to move to subject position (i.e. SPEC of IP position), where it is assigned Case by INFL; this g-Case indexing results in subject/verb agreement.



Otherwise, if the verb assigns Case, the complement NP can remain inside the VP; Move α need not apply for the NP to be in a position where it can be assigned Case.



In these circumstances, the statement, in X-bar theory, that complements are to the right is redundant. If the verb has Case to assign to its complement and it assigns Case to the right, the complement will have to follow the verb and get Case there. If the verb has no Case to assign, the complement will have to move outside the VP. Either way, it is unnecessary for X-bar theory to say that the complement is to the right of the verb⁶.

Equally, it would be possible to suppress information about the direction of Case assignment and specify only in X-bar theory the order of complements and specifiers with respect to heads. In fact, this solution would be preferable to the previous one because it can generalise over the whole X-bar hierarchy, while Case theory constrains word order only in so far as head categories are Case assigners. Thus we could give a relative order of heads, modifiers and specifiers, and each type of head assigns Case (if it does) in the direction in which its NP sister appears canonically.

Motivation for keeping both notions of directionality would arise if it made sense to think of a language where, say, the direction of Case assignment from the verb to

⁶In the case of subjects, matters are in fact more complex, since, in Catalan, there are subjects which follow the verb, as in *Arriben els nens*, literally 'arrive the boys', 'the boys arrive' or *Salten les nenes*, 'jump the girls', 'the girls jump'. So, subjects of intransitive verbs can be either to the right or to the left of the inflected verb.

the object is to the right and, at the same time, the complement of the verb is canonically generated to the left of the verb. But, in those circumstances, there is no reason to base-generate the complement to the left, to have it always moving rightwards, unless there is evidence of NP movement. In general, it would seem advisable to suppress one of the two independent statements about directionality.

However, Travis (1989) points out the complexity of word orders found in some languages, and proposes that this can be handled by word order parameters in both X-bar theory, and Case theory or theta theory, of which the X-bar theory directionality parameter applies as a default. She motivates this approach with data from Chinese and Kpelle. In Chinese, theta-marking is to the right (so that directly theta-marked complements appear to the right of the head), while the value for the X-bar parameter is head-final (so that complements appear to the left of the head if not directly theta-marked by it). In Kpelle, on the other hand, directionality of Case assignment (instead of theta-marking) is parameterised, and there is also an X-bar parameter of headness.

Augmenting the number of specifications of directionality in the grammar seems necessary in view of the array of word orders found in natural language. Accordingly, in what follows I assume that, in Catalan, directionality of Case assignment takes care of, for instance, the order of a verb or preposition and its complements, while also there are X-bar theoretic default specifications of order for situations unspecified by Case theory. For non-Case marked constituents, the default X-bar directionality parameter is decisive.

4.4.1. *X-bar theory and word formation*

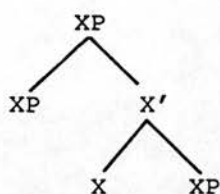
The X-bar theory outlined, for example, in Stowell 1981 and Chomsky 1970 does not include order of constituents inside a word. So there is the question of how to determine, in a grammar, the order of word-internal constituents. One can avoid the issue for affixation because each affix is marked as preceding a word (if it is a prefix) or as following it (if it is a suffix), on a completely idiosyncratic basis. The question cannot be avoided for compounding, though, where the order $[X^0 Y^0]$ (X^0 , Y^0 being non-affixal) is possibly grammatical, whereas $[Y^0 X^0]$ is not.

What are the grammatical properties that could possibly determine the order of the constituents of a word? One cannot assume that the order of those elements is the same as it is sententially (as the Catalan data shows); but, can the criteria to determine it be the same? Sentential word order is fixed by the parameterisation of X-bar theory and Case theory, and there is no reason to think that this does not also apply below the word level; indeed X-bar theory has already been widely

applied to morphology (cf. Selkirk 1982 and Sadock 1985, including "negative" X-bar levels).

Let us examine the possibility of fixing the order of constituents in a compound by the parameterisation of X-bar theory. Naturally, this parameterisation can only make reference to X-bar concepts such as head, complement and specifier, and it is not clear that the distinction between complement and specifier can be maintained within the X^0 level. Given the basic structure in (39), the XP which is a sister of X' is its specifier; the XP which is the sister of X is its complement. One can only tell apart specifiers and complements by reference to X' vs. X. If only X^0 categories are posited in compounding, the distinction specifier/complement cannot be maintained.

(39)



(order irrelevant here)

So, let us suppose, minimally, that the notions of head and non-head are sufficient. It has been proposed that this is enough to determine the order of the constituents of words in English: Williams' (1981b) Righthand Head Rule (RHR) defines "the head of a morphologically complex word to be the righthand member of that word" (though there are exceptions to the RHR, such as the prefix *en-* which is the head of *enslave*, etc.). This kind of statement has its place in a GB grammar as part of X-bar theory when referring to its parameterisations at the word level. Catalan compounds are also commonly head final:

- (40) $[[\text{curt}]_A [\text{circuit}]_N]_N$
 short circuit
 $[[\text{cor}]_N [\text{trenca}]_V]_V$
 heart break

The only exception is found in the *parallamps* type compounds:

- (41) $[[\text{sota}]_P [\text{escala}]_N]_N$
 under staircase
 'space under the stairs'
 $[[\text{para}]_V [\text{brises}]_N]_N$
 stop breezes
 'wind screen'

The X-bar schemata are limited in their scope to headed constructions; in Williams 1981b, as noted, exocentric words are considered headless and generated by unary

rules, so that, for Williams, these require an independent treatment. In my analysis of the *parallamps* type, these compounds are considered headed (though still exceptional, because unlike all other compounds they are *left-* and null-headed).

4.4.2. Feature Percolation Conventions

Lieber (1980, 1983) proposes feature percolation conventions which have the same function in the grammar as Williams' RHR. These conventions are adopted here, with minor modifications, for the description of trees representing the structure of words. They are assumed to apply on word structures which are trees limited to unary and binary branching⁷. The leaves are lexical items, and Convention I simply states that a lexical item is dominated by a node with its features.

- (42) Convention I
All features of a stem morpheme percolate to the first nonbranching node dominating that morpheme.

This is illustrated in (43):

- (43)
- | | |
|------|-------|
| N | A |
| | |
| leaf | green |

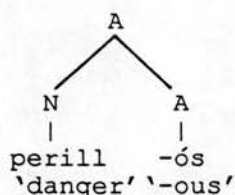
I shall maintain Lieber's original terminology in referring to stems, because it is understood that the distinction stem/word may need to be made in the grammar. In particular, for the purposes of affixation, it has been argued that some affixes attach to stems and some others to words, the distinction between the two being phonological (cf. Sproat 1985:459ff). In any case, following Selkirk 1980 and Lieber 1980, 1983, I take compounding to be the concatenation of words, i.e. non-affixal X^0 projections. So, for present purposes, stems may be equated with simple words.

- (44) Convention II
All features of an affix morpheme percolate to the first branching node dominating that morpheme.

This is illustrated in (45).

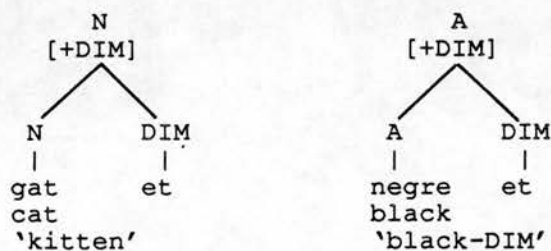
⁷The model proposed here has been devised with a concatenational system in mind - and this is what is in fact required for compounding. Proposals have been made to deal with more complex systems (e.g. McCarthy 1981 on discontinuous morphemes in the Semitic languages), but I shall not enter this matter here.

(45)



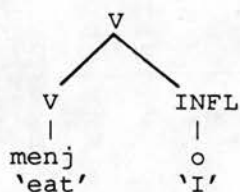
Some affixes (e.g. the diminutive marker) have no category features, and so, for instance, an adjective followed by a diminutive marker is an adjective, while a noun followed by the diminutive is still a noun, as in (46).

(46)

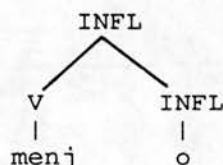


Convention II is the first convention that can determine the syntactic category feature of a node. In this connection, a conflict seems to arise in the case of I (or INFL) between syntactically driven descriptions of word constituency and the morphologically driven ones. On the one hand, the inflected verbal form is taken to be a verb in morphology, as in (47a). On the other hand, the movement of V to INFL in the syntax is usually taken to give the structure in (47b) (assuming that head movement involves adjunction), though the resulting category is often symbolised by V_I .

(47) a.



b.



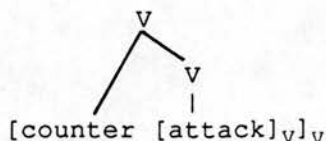
There are nodes which fail to be given a value for certain features by Convention II; then Convention III applies:

(48) Convention III

If a branching node fails to obtain all features necessarily associated with its category or fails to obtain features at all by Convention II, features from the next lowest node automatically percolate up to the underdefined branching node.

E.g. (49).

(49)

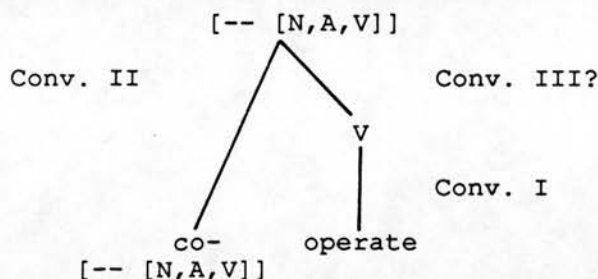


Originally, this convention was phrased (in Lieber 1983:253) as follows:

"If a branching node fails to obtain features by Convention II, features from the next lowest labelled node automatically percolate up to the unlabelled branching node."

My modification responds to a criticism of Namiki (1982). It is argued that, by Lieber's conventions, in a word like *cooperate*, it would be impossible to know which category the word belongs to. Consider (50).

(50)



Namiki assumes that subcategorisation features percolate, so let us assume this for the sake of the argument. If Convention III applies only when the features of an affix fail to percolate, Convention III does not apply in (50). Therefore, the word fails to have a category. But it is obvious that subcategorisation features of the affix should not percolate to the first branching node, because the subcategorisation features of the affix tell what the requirements of the affix are towards its sister. In general, however, Namiki is pointing out that, in Lieber's formulation, the application of Convention II precludes the application of Convention III, and that this may give the wrong results. Altogether, it seems that the mother branching node can get its features from more than one of its daughters. This is what has been termed "relative headness" by Di Sciullo and Williams (1987), under the assumption that heads of a construction are the constituents which determine its features (syntactic category and others, to be specified). (The idea of "relative heads" will be employed in what follows). This is what this minor reformulation of Convention III is aimed at.

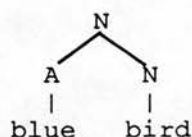
The final convention is applicable to compounds:

(51) Convention IV

If two stems are sisters (i.e. they form a compound), features from the right-hand stem percolate up to the branching node dominating the stems.

This convention is language-specific and initially formulated for the compounds of English, as typified by (52).

(52)



If the position of the head in a compound can be parameterised, left-headed compounds are in principle allowed. This relates to the controversy between "asymmetric" and "symmetric" proposals in morphology (cf. Namiki 1982). As opposed to the "asymmetric" proposal, Lieber's "symmetric" one allows for words to be either left- or right-headed. Lieber claims that there are languages such as Vietnamese and Thai, with head initial compounds. Namiki's (1982) and Trommelen and Zonnevelt's (1986) arguments against such a statement are not conclusive, because they rest on the assumption that head initial compounds are marked, meaning, as far as I can see, rare, i.e. statistically infrequent. But this does not mean that they should be excluded in a grammatical characterisation. Overall, Namiki's and Trommelen and Zonnevelt's arguments appear to be an attempt to salvage the RHR of Williams 1981b by dismissing a counterexample. Worse than this, the "asymmetric" proposal does not seem preferable from a theoretical perspective. An "asymmetric" model is simpler than a "symmetric" one in that it allows one to account for the relative order of head and non-head in morphology in general with one statement only, but this simplicity is not made substantive with any kind of explanation. It is comparable to a phrase structure grammar which states that in all languages VPs appear in a particular position (and manages to reduce all "apparent" counterexamples to this regularity), without making any attempt to explain why⁸.

Like Lieber, and unlike Williams 1981b, Baker (1988b) presents evidence in favour of an independent convention for the percolation of features in compounds as opposed to other complex words. Baker's evidence is taken from Tzeltal, a Mayan language, in which all affixes are suffixes, while compounds are head initial:

- (53) pih 'intelligent'
 pih-il 'intelligence'
- hol-wič
 head mountain 'summit' (Tzeltal)

Head initial and head final compounds may coexist in one language; this is what Scalise (1988) advocates for Somali (where [N N] compounds are head initial and [N V] ones are head final), and for Italian. Though Scalise's examples from Italian

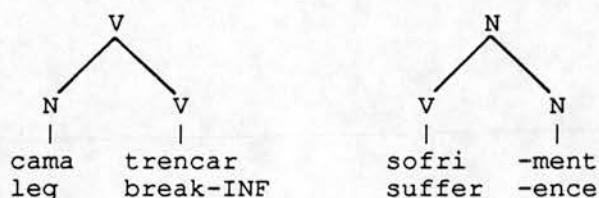
⁸One fact that the "symmetrical" model has been unable to account for is the relative frequency of suffixes over prefixes in natural language. Notwithstanding, all that the model of morphology referred to here addresses is the question of what is possible, not what is frequent in word formation. It does not come as a surprise, then, that the issue of frequency has not been solved.

are not very convincing in that the position of the head of a compound seems to depend on the historical period of the compound's formation (Latin compounds seem to be head final, and modern Italian compounds head initial), the data from Somali seem quite robust.

So, on empirical grounds, Convention IV can be adopted for English, but, when dealing with other languages, another convention might have to be adopted. This possible alternation has the status of a parameter. In the simplest case, the two values of it are "right" and "left" in the headness of compounds, and the value for the parameter in English is "right". For authors like Boase-Beier and Toman (1987), Case theory can determine the order of constituents inside complex words; if so, the parameter established in Convention IV can be thought of as a Case theoretic parameter.

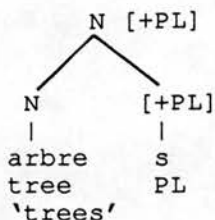
If one takes the notion of relative head seriously, differences and similarities arise between heads in morphology and heads in phrases. In both contexts, heads designate constituents whose features percolate to the mother node. There is a division between features that percolate and those which do not. The first will be called head features; the second are irrelevant at the level of the mother node. In the context of X-bar syntax, a phrase XP is a projection of X, where X stands for a syntactic category; so, syntactic category is the relevant head feature for phrase well-formedness. Syntactic category is also a head feature in word structures, illustrated in (54).

(54)



Yet, there are other features which must percolate to the mother node, to specify the behaviour of the composite in a phrasal context. If not all syntactically relevant features appear on the same daughter, features must percolate from more than one daughter (and all of these daughter constituents are heads relative to one or several features). E.g. in (55) there is a constituent for the feature [+PL], which must percolate for the complex word to be marked as plural. Also, syntactic category is a necessary feature of a word and, therefore, must percolate from (another) daughter constituent. The two constituents characterise the word as a whole.

(55)



In Zwicky 1985 the argument is made that, far from simplifying and/or improving the power of prediction of the grammar, heads in morphology are not particularly useful, except for certain types of compounding. This much is consistent with the extensions discussed above. However, he goes on to argue that heads in phrases and in words are not analogous, and this is a position for which some criticism can be raised⁹.

Zwicky's argument goes as follows: (i) He claims that the relevant notion for headness in syntax is that of morphosyntactic locus (a notion independent of all others, e.g. semantic functor, semantic argument). (ii) When considering word structure, not only the notion of morphological locus emerges as a head-like notion, but also the notion of morphological determinant (morphological determinants are the constituents of a word which determine its morphosyntactic features). Moreover, the locus of inflection seems to be determined more by margins of words than by heads. (iii) Zwicky claims that syntactic determinants (i.e. the constituents which share the syntactic category of the whole) could not play a role in the determination of heads in the syntax. This is so because syntactic determinants do not coincide with morphosyntactic loci (in syntax). The conjunction of (i) to (iii) leads Zwicky to the conclusion that there is no parallel between heads in syntax and heads in morphology (in fact, that heads in morphology are practically dispensable). Nevertheless, if (iii) were challenged, this conclusion would not be reached; that is, Zwicky does not seriously discuss the notion of syntactic determinant as phrasal head because it does not coincide with that of morphosyntactic locus, which he has already identified with the notion of head in phrases.

Thus, it seems possible to propose, in opposition to (i) to (iii) above, that: (i') Heads in syntax correspond to syntactic determinants¹⁰. (ii') Heads in morphology correspond to morphological determinants (and headless, exocentric, words -- if they can be found -- fall outside the scope of any consideration on heads, and

⁹Recently, Bauer (1990) has also argued for the notion of head in word formation in response to Zwicky's 1985 paper; his argument is different from mine, though leading to a similar conclusion.

¹⁰These are, for the six structures of English considered in Zwicky 1985, the following: P in [P NP]; V in [V NP]; VP in [NP VP]; AUX in [AUX VP]; COMP in [COMP S]; and DET in [DET N]. Except for the last one (which is controversial) and for [NP VP], which does not apply -- because subjects are sisters of I' and not VP in GB -- these choices correspond to heads as usually considered in this framework.

deserve attention elsewhere). (iii') Therefore, syntactic and morphological determinants are heads in syntax and morphology respectively. As a consequence, the notion of head basically coincides in one domain and the other, hence the parallelism phrase structure/word structure holds. In actual fact, it is possible to argue that the second is a prolongation of the first within the X^0 projection.

In conclusion, I shall consider that (relative) headness (and, with it, percolation) has as its purpose the morphosyntactic determination of linguistic constructions.

Some inherent syntactic features particularly relevant in this thesis are shortly considered next. As a general point, note that features of words need not be universal; while some are commonly agreed to be universal (for instance, syntactic category, and Case and theta theoretic features), others are considered language-specific; for example, Borer (1988) considers definiteness a head feature in Hebrew for all nominal categories. The set of values that a feature can take can also vary cross-linguistically (e.g. that number is a universal feature does not imply that the set of realised values for the feature number is the same in all languages).

1. Syntactic category

It has been assumed that syntactic category (often called lexical category) is determined by two binary features, $[+/-N]$ and $[+/-V]$; the usual denominations N, A, V and P are used for short, standing for $[+N, -V]$, $[+N, +V]$, $[-N, +V]$ and $[-N, -V]$ respectively (cf. Chomsky 1970). Constituents such as DET are determined by specific features like $[+DET]$.

Once syntactic category has been specified, the complex word must also be characterised with respect to morphosyntactic features -- features consistent with the syntactic category (e.g. generally verbs cannot be specified for gender). (Some of) the facts which determine the features of a word then are:

- Of all the features F for which a word can be characterised, each category is characterised by a subset of features: F_N (features of a noun category), F_V , F_A , etc. (cf. Bach 1983). This correlation is partly language specific (e.g. not all languages mark their adjectives for gender), but the acquisition of this variation is presumably facilitated by morphological evidence. E.g. in Catalan¹¹,

¹¹In Catalan, only pronouns have morphological case, so this is a feature of NPs, not nouns. Also notice that the features under F_A are not inherent, but rather acquired by adjectives under agreement.

(56) F_N : gender {MASC, FEM}
 number {SING, PL}

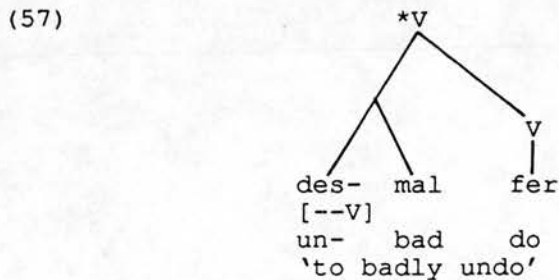
F_A : gender {MASC, FEM}
 number {SING, PL}

F_V : tense {PRES, PAST, FUT}
 mood {IND, SUBJ, IMP}

- For each feature for which an element can be specified, one and only one value can be assigned; e.g. a noun cannot be assigned at the same time two feature values for gender, compatible among themselves or not. So, the feature values which define a word must be consistent among themselves and fully characterise the word given its syntactic category¹².
- F_N , F_A , F_V , F_P are a subset of the features (in a broad sense) contained in the lexical entries of N, A, V, P, respectively. (For lexical entries, see section 3.1.1.).

2. Subcategorisation features

Notwithstanding what has been decided for non-affixal categories (see 3.1.1.), it will be assumed, as in standard practice, that affixes have subcategorisation frames in their lexical entries; the subcategorisation frame of an affix is one of the features which does not percolate to the mother node. That is, subcategorisation requirements must be satisfied by the sister node of an affix, as shown by the ungrammaticality of the example in (57).



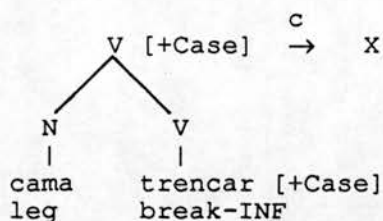
It remains a topic for future research to see if these subcategorisation frames could be dispensed with for affixes -- in favour of theta or Case theoretic specifications, and whether this move would be motivated.

¹²As a result, underspecified trees at S-structure are ruled out. Thus *sheep* must be specified, at S-structure, as [-PL] or [+PL], and it is not allowed without a specification. Ambiguity results from the possibility of having more than one specification and, consequently, more than one interpretation. So, only a fully specified category can label a node; this does not stop us from generalising over different kinds of elements.

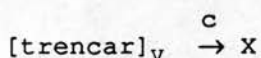
3. Case features

Case features are head features, in compounds of Catalan, as has already been shown; e.g. (58).

(58) a.



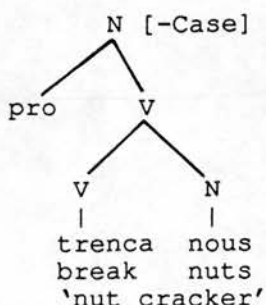
b.



(The arrow marked with a "c" indicates Case assignment, so that a constituent X is assigned Case equally by *trençar* and *camatrençar*). If Case features are head features, the CFPP of Baker ((12) above) follows. That is, a complex word inherits the Case assigning properties of the head only; in consequence it cannot have a Case assigning capacity bigger than that of the simple elements which constitute it -- indeed than one which is of the same category as the complex word, since syntactic category is a head feature as well.

What seems peculiar is the following, exemplifying the *parallamps* type.

(59)

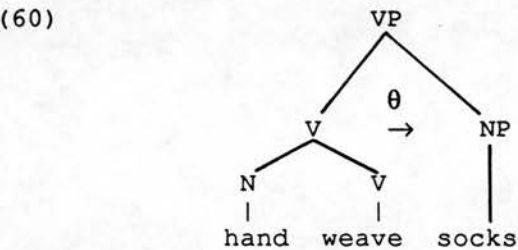


Here *trençanous* as a whole has the Case theoretic properties of its head, pro_N ; this e.c., like all nouns, cannot Case-mark any complement. However, *trenca* does not seem to satisfy its Case assigning capacity, if -- as assumed for all compounds -- *nous* is not assigned Case. For the analysis of the *alatrençar* type I have assumed as well that a Case assigner must assign Case. A way of reconciling all these facts involves the following hypothesis: a verb can only realise its Case assigning capacity when it has adjoined to

INFL (i.e. INFL licenses the assignment of Case by the verb)¹³. Thus, *trenca* in *trencauous* does not have any Case to assign because it has not adjoined to INFL; this is the difference, then, between *trenca* in *trencauous* and in *alatrencau*, as far as Case theoretic properties are concerned. In adopting this hypothesis, all previous assumptions on Case can be maintained.

4. Thematic features

These have been claimed to be head features, and so to percolate, according to Lieber's (1983) analysis of English compounds:



However, they cannot be claimed to be head features universally (see section 5.2. and Chapter Six for Catalan compounds and compounds in some dialects of English), and therefore some parameterisation seems required.

4.4.3. Some speculations on word order

Let us come back now to the more general issue of the relationship between word-internal order of elements -- in particular in compounds -- and phrasal word order.

Suppose we have a theory of grammar which, taking the Strong Lexicalist Hypothesis to one of its extreme conclusions, postulates independent principles for the determination of word and sentence well-formedness. In such a theory, in the case that the order of the constituents of a compound and the order of those elements at the phrasal level is the same, this fact has to be taken as accidental. In fact, it is hard to see in what way such a theory can account for any characteristic that might generalise over words and sentences. Accepting, for instance, the X-bar convention at all levels constitutes a weakening of the position outlined: absolute independence between the word level and the level above it.

On the other hand, a theory of grammar based on the assumption that there is no

¹³Gerundive affixes and non-finite verbal forms like them may also license Case assignment. This is a topic for future research.

difference between word formation and phrase generation has no way of explaining -- or even describing -- the discrepancies in constituent order inside and outside words. That is, this kind of theory predicts the same constituent order inside and outside words. There is evidence against this, and also evidence which points to the fact that certain words are not subject to particular (syntactic) principles (see, for instance, Borer's comments on the PrPr, reported in section 3.1.2.2.).

In the principles-and-parameters treatment of word formation that I am outlining, neither of the two above positions is given support. Both parameters and principles are going to apply across the board if their conditions of application hold, but there is no prohibition on principles stated as applying only to certain word structures or phrase structures. No generalisations which hold for words and phrases need be lost.

Given this quite flexible framework, one still has to account for the fact that word internal order (expressed in terms of heads) sometimes coincides with sentential word order, and sometimes is precisely the opposite of it. In fact, some languages seem to pattern one way or the other consistently. In Catalan, the internal order of compounds is the opposite of sentential word order (with the exception of the *parallamps* type). Like Catalan compounds, Fijian compounds display an order of constituents which is the inverse of that found in phrases (Vernon Shilliday, p.c.).

- (61) a. [prim filar]_v
 fine spin-INF 'to split hairs'
 *[fila prim]_v
 [filar prim]_{vp}
 spin-INF fine 'to split hairs'
 or 'to spin finely'
 *[prim filar]_{vp}
- b. [ala trencar]_v
 wing break-INF 'to break some-one/
 thing's wing/s'
 *[trenca ala]_v
 [trencar una ala]_{vp}
 break-INF a wing 'to break a wing'
 *[una ala trencar]_{vp}

So, in Catalan and in Fijian there is complementarity between phrasal word order and word order inside compounds. Such an asymmetry could be explained as a means to facilitate processing, that is to identify one structure as a phrase and the other as a word. We would have a functional explanation for the existence of opposite word orders in phrases and compounds.

What happens then with other languages (such as English), in which the order of words in a compound is the same as the phrasal word order (e.g. *blackbird* vs. *black bird*)? The function served by the asymmetry in the Catalan and Fijian examples would have to be fulfilled by some other means. In English, for example, it appears to be accomplished by stress patterns. Another way of marking the difference between words and phrases is the position of morphological markers.

There are various ways for a language to express the difference between words and phrases, and some coexist in the same language. For example, it has been shown that Catalan compounds are different from phrases with regard to morphology and Case assignment, as well as word order. (Notice that, in any case, it is not possible to appeal to semantic differences between words and phrases because, ultimately, this is our departure point. That is, some difference in meaning -- possibly that between referential and generic -- is what is expressed in the distinction word/phrase).

The tendency to have asymmetries between words and phrases is, in this context, understandable. However, this is more descriptive than explanatory: given the tendency, it is still impossible (in the absence of more information) to know whether the distinction word/phrase is going to be preserved by this or that other means. As it stands, this explanation has no predictive power. Perhaps, the means used in each particular language correlate with other parts of the grammar, and, once the correlation had been established, it would be possible to start predicting which form of grammaticalisation will be used by a given language.

It is possible to establish a principle requiring the distinction compound/phrase to be materialised in some way or another. The distinction must only necessarily be imposed, among words, on compounds, since they are the only words which can possibly be confused with phrases (because their constituents are words themselves). Like the Principle of PF Interpretation ((10) above), this is also a principle of PF interpretation, i.e. a principle requiring a grammatical distinction to be kept at PF.

- (62) Principle of PF Interpretation of compounds
The distinction between a compound and X'/XP
projections must be interpretable by the rules of PF.

This principle can be read as a constraint on the construction of a theoretical grammar: no systematic distinction between structures can be postulated that does not manifest itself in PF.

Having established the possible complementarity between word order inside compounds and at the sentential level because of functional reasons, one might still wonder if it is possible to account for word order in terms of other notions.

Assuming binary branching, the problem of word order is reduced to determining whether constituents are to the right or to the left of each other. There have been quite a number of attempts to relate phrasal order to other primitives belonging to information theory, or processing. (For an application of information theory to Catalan, see Vallduví 1988).

For complex words, several proposals could be advanced, trying to account for word order inside complex words on the basis of (i) adjacency of heads to inflection; (ii) inflection being adjacent to the category type that it modifies; (iii) constraints on the operation of adjunction. As we shall see, none of these accounts is entirely satisfactory.

Among complex words, take compounds with peripheral inflection. For languages such as English and Catalan, which have right-headed compounds, the situation schematically represented in (63) holds:

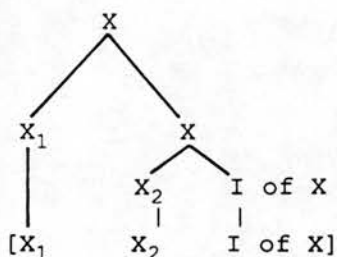
- (63) a. X H I
 b. *H X I
 c. *I X H
 d. *I H X

H stands for "head", X for its complement or modifier, and I for "inflectional marker".

According to (i) above, the head and all inflectional markers must be adjacent, thus (63b) and (63c) are ruled out. Because inflectional markers happen to be suffixes (and not prefixes), (63d) can also be ruled out. This makes the right predictions for English and Catalan, since (63a) is the only sequence which is grammatical. Yet this approach predicts that all languages with head-initial compounds present prefixed inflection, which is not the case. For example, this hypothesis is falsified by the case of Tzeltal ((53) above), where all affixes are suffixed, while compounds are head initial.

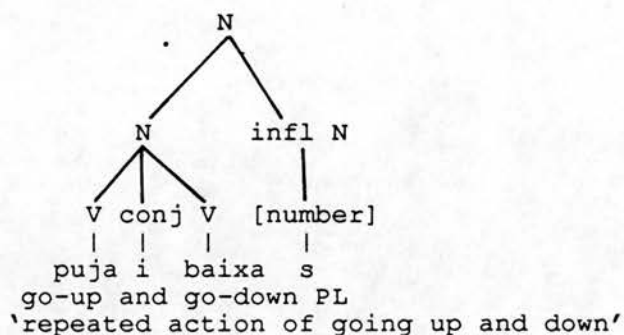
The second hypothesis, (ii), is based on the same assumption as (i), namely that inflection is peripheral and, in Catalan, suffixed. Hypothesis (ii) imposes a weaker condition on the attachment of inflection: it must attach adjacent to the category type it modifies, but not necessarily to the head of the word. This allows the generation of the well-formed compounds which are found in Catalan, but also of head-initial compounds of the form:

(64)



where X_1 is the head of the compound, and, accidentally, of the same category as X_2 . These are the circumstances in which the inflection of X attaches to the non-head. This hypothesis can also be dismissed because it wrongly predicts that (65) is ungrammatical (because of a nominal inflectional marker being adjacent to a verb); (65), though rare, is grammatical.

(65)



Finally, (iii) only rephrases the problem. To the question: Why are English and Catalan compounds right-headed, it answers: Because they are the result of left adjunction of a complement or modifier to a head. Left adjunction stands for right-headedness, and therefore (iii) is far from explanatory.

Given the shortcomings of (i)-(iii), it seems clear that (right/left) headedness of compounds in different languages must be explicitly stated in the grammar; for the time being, it remains underived from other linguistic phenomena.

Chapter Five

Government theory and categorial conditions

5.1. Government theory and categorial constraints

In this section, after considering the notion of government, I present a problem dealt with on the basis of this notion in the literature; I proceed then to consider alternative solutions.

5.1.1. *Government and word formation*

5.1.1.1. *On the notion of government*

Central to the theory of Government and Binding is the notion of *government*, partly inherited from previous notions of traditional grammar, but new in its specificity and explicitness. Government is defined in the following manner in Chomsky 1986b, which I shall take as the basis for the discussion:

- (1) α governs β
iff α m-commands β and
there is no γ , γ a barrier for β ,
such that γ excludes α .
(Chomsky 1986b:9)

This definition rests on the notions of m-command, barrier, exclusion which are reproduced below.

- (2) α c-commands β
 iff α does not dominate β and
 every γ that dominates α dominates β .

Where γ is restricted to maximal projections (...),
 we will say that α m-commands β .
 (op. cit.:8)

It is additionally required that in (2) β does not dominate α . This stops, e.g., c-command of a category by itself (see op. cit.:92, note 12)).

That α and β exclude each other means that they have no part in common:

- (3) α excludes β if no segment of α dominates β .
 (op. cit.:9)

Barrier and related notions are as follows:

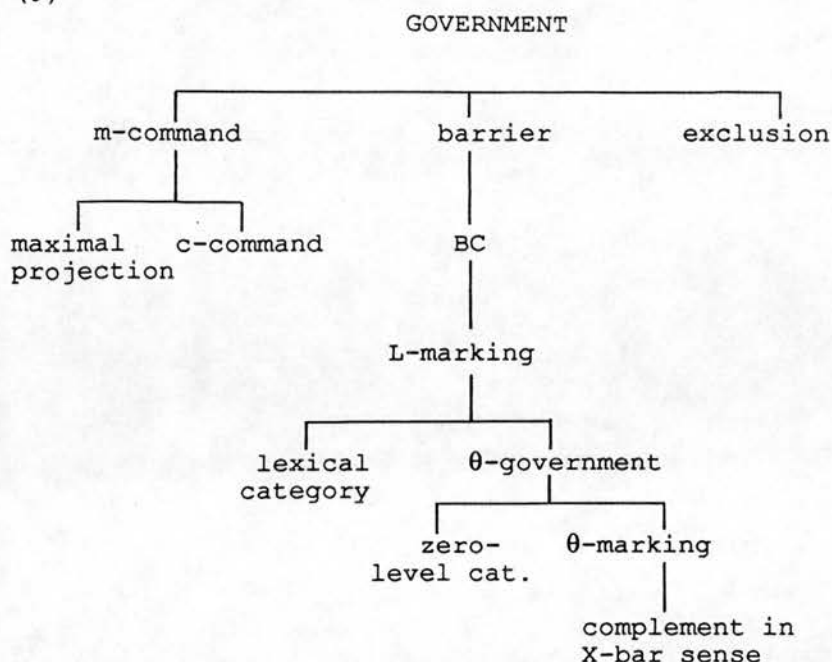
- (4) γ is a BC [blocking category]
 iff γ is not L-marked and γ dominates β .
 (5) γ is a barrier for β iff (a) or (b):
 a. γ immediately dominates δ , δ a BC for β ;
 b. γ is a BC for β , $\gamma \neq \text{IP}$.
 (6) α L-marks β
 iff α is a lexical category that θ -governs β .
 (7) α θ -governs β
 iff α is a zero-level category that θ -marks β ,
 and α , β are sisters.
 (op. cit.:14-15)

The following condition holds on direct theta-marking:

- (8) α directly θ -marks β only if
 α and β are sisters.

The definition of government in (1) rests on the notion of barrier and hence ultimately on X-bar theory (domination, complement, sisterhood, zero-level category, maximal projection) and the notion of lexical category (here meant to include nouns, verbs, prepositions and adjectives). See (9), where the lines link notions interrelated in their definitions: e.g. the notion of government depends on the notions of m-command, barrier and exclusion, etc.

(9)



The concept of barrier is essential to consider structures in which movement plays a role. A second notion of barrierhood is defined by the Minimality Condition (op. cit.:Section 8); this will be the one relevant for the treatment of compounding in this chapter.

- (10) γ is a barrier for β
if γ is (a projection, the immediate projection) of δ ,
a zero-level category distinct from β
in the configuration
 $.. \alpha .. [\gamma ... \delta ... \beta ..]$
(op. cit.:42)

The Minimality Condition is meant to express the intuition that "the complement of a head cannot be governed by a more remote head" (op. cit.:43).

The choice between "a projection" and "the immediate projection" in its formulation corresponds, respectively, to a broader and a narrower formulation of barrier; the narrower concept is preferred on the evidence of Romance extraction of subject of NP (see op. cit.:44ff) and other phenomena, and so it will be the "immediate projection" version that I shall adopt here.

Empty categories interact in a particular way with Minimality: "in [(11b)] *t* is protected from antecedent government by *C* (=that), by virtue of the Minimality Condition, but in [(11a)] this will not be the case if we make the natural assumption that *e* is "featureless" and therefore does not serve as an appropriate choice for δ in [(11)]; the intuition is that a minimal governor must be a category

with features to serve as a barrier to government" (op. cit.:47). Throughout this discussion, "featureless" means without Φ -features, i.e. the features person, number, gender, etc.¹ (see op. cit.:25 and Chomsky 1981).

- (11) a. who did you believe [_{CP} t' [_C e [_{IP} t would win]]]
 b. *who did you believe [_{CP} t' [_C that [_{IP} t would win]]]

Being central in GB syntax, the notion of government (in a broad sense) is used, perhaps more tentatively, in other components of the grammar, e.g. the phonology, in the theory of charm and government (cf. Kaye et al. 1985). The spirit of this approach is expressed in the following way (Kaye et al. 1985:314):

"A phonological sequence is not merely a string of autonomous syllables. Numerous phonological processes serve to bind syllables of a given domain (word, clause, sentence) together. Stress, tone, harmony, assimilation, etc., may all be viewed as serving this function in one way or another. What these phenomena have in common is the presence of a dominant unit of some form in relation with one or more subordinate units. We call this relation GOVERNMENT. It is our view that charm is an agent through which government is expressed".

In what follows, I shall look at a possible use of government in the domain of X^0 projections, in particular the domain of complex words.

5.1.1.2. A problem

As far back as Di Sciullo 1981, the notion of government has been used in this framework to constrain the structure of compounds. In Di Sciullo's paper, the examples in (12) below are ruled out because no government relation holds between the verb and the element that precedes it (COMP, Adv, or external argument NP).

- (12) a. *qui-trompe *COMP V
 who deceives
- *à-quoi-pense
 in what thinks
- b. *l'oeil-trompe *NP_{SUBJ} V
 the eye deceives
- *la-soif-coupe
 the thirst cut

¹Or, rather, with no values for the features number, gender, etc.

c. *nécessairement-boit
necessarily drinks

*Adv V

*boit-nécessairement

According to Di Sciullo and Williams' (1987:85) summary of Di Sciullo 1981

"The structure in ([12a]) is excluded because the V does not govern the COMP position. ([12c]) is excluded because V does not govern sentence adverbs such as *nécessairement*, and ([12b]) is excluded because no government relation holds between the V and its external argument."

The principle to account for this is as formulated in (13):

- (13) Dans une structure $[[V] [X^n]]_N$
où X est une catégorie lexicale, V doit pouvoir
gouverner X^n au niveau des structures syntaxiques
de base.
(Di Sciullo 1981:57)

In connection with this approach, two questions arise:

1. At what level of representation is the (missing) government relationship supposed to hold? It seems that the argument is that *COMP V is out because in a phrase structure the verb would not govern the COMP position (and likewise for adverbials and external arguments). That is, the examples in (12) are ill-formed because of the ill-formedness of configurations which are closely related to them. The government relation holds, then, in configurations other than those underlying the examples in (12).
2. With respect to (12c), Di Sciullo argues that, since *nécessairement* is a sentential adverb, the verb does not govern it. Now the question is whether any adverbs, sentential or not sentential, are allowed to compound with verbs, as (14) seems to suggest:

- (14) mal-traiter
badly treat 'to abuse'
- mal-mener
badly lead 'to handle badly'

There is the possibility that the words in (14) are not compounds and, hence, irrelevant for Di Sciullo's (1981) account, but I see no a priori reason to exclude them as such.

In the light of (14), it is natural to relate the ungrammaticality of (12c) to the fact that it contains an adverb in *-ment*. I shall examine a bit more closely the adverbs in *-ment* in French, *-ment* in Catalan, and *-ly* in English, which seem to share

characteristics across these languages. However, the solution that I propose for the problem in [2] above will not rest on the use of the notion of government. On the other hand, I shall address the problem in [1] in the first part of this chapter and present a use of the notion of government in word formation alternative to the one assumed by Di Sciullo (1981).

5.1.1.3. Government theory and compounding

The theory of government may be expected to play a part in the description of well-formedness in compounding, given the basic rationale behind government theory, namely that government is a structural relation holding primarily between an item and its complement, e.g. a verb and its internal argument. Government holds between a Case assigner and the argument that it assigns Case to.

It has been argued above that there are compound internal theta role assignments -- exemplified here by *alatrencar* 'to break some-one/thing's wing/s'. If this is so, and theta role assignment occurs under government, it follows that the relation between the constituents of *alatrencar* is one of government. Specifically, *ala* 'wing' is governed by *trencar* 'to break', in the same way that *trencar* governs *la cadira* 'the chair' in (15).

- (15) no volia [trencar [la cadira]_{NP}]_{VP}
 no I-wanted break-INF the chair
 'I didn't want to break the chair'

Nonetheless, the governed constituent in (15), *la cadira*, is a maximal projection; not so *ala* in the compound. The global issue that concerns us here is whether, as initially formulated, the various principles of the subtheories of GB can be applied below the X^0 level or need to be modified to be applicable. Theta-marking, which is one of the notions underlying the definition of government, is usually assumed to apply to a maximal projection; I assume here, though, that a X^0 projection can also be theta-marked.

Hence, contrary to the proposals of Di Sciullo in her 1981 paper, I understand government to hold in a compound, and thus to license compounds on the basis of their own internal structure, not on the basis of some other structure. Of course, the use of the notion of government made here will be quite simple because the structures dealt with are all of the same form, namely $[X^0, Y^0]_{X^0}$ ².

Consider the following:

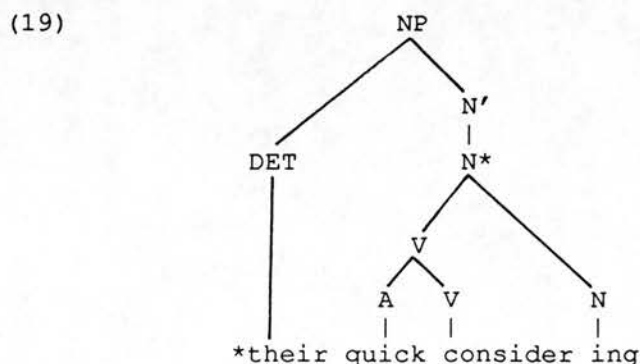
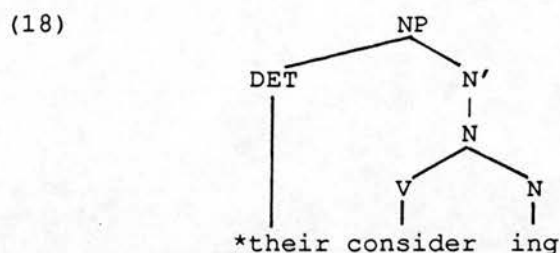
²The comma indicates that the order of X^0 and Y^0 is unspecified here.

- (16) a. they are [considering]_V the proposal
 b. *their [considering]_N went on for a week

Lieber (1983) presents the contrasts in (17). These will be handled in terms of government.

- (17) a. they are [quick-considering]_V the proposal
 b. *they are [proposal-considering]_V the offer
 c. *their [quick-considering]_N went on for a week
 d. their [proposal-considering]_N went on for a week

In Lieber's system: (i) Thematic features are head features, and (ii) Identity of syntactic category is a condition for the percolation of other features. Therefore, when [X *considering*] is a noun, the argument of the verb *consider* must be satisfied inside the compound (in order for it not to violate the Argument Linking Principle) because of constraints in percolation. The claim here is that the category changing affix introduces a new kind of head for a structure and, as such, operates as a node blocking government of arguments. This is the reason why (16b) and (17c) are out.

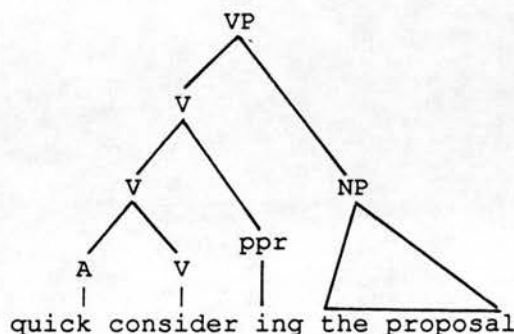


In (19) *consider* does not have its internal argument satisfied, unlike (17d), which is grammatical. Notice that the configuration in (19) is identical to the one that the Minimality Condition refers to: N* is a barrier for V if N* is the immediate projection of N, a zero-level category distinct from V, as in the configuration in (20).

(20) ... α ... [_{N*} ...V...N...]

On the other hand, there is no barrier in the case of [X considering] as a verb, because *considering* and the verb *consider* are of the same syntactic category, thus all features of *consider* must be inherited by *considering*, and *-ing_V* does not block the satisfaction of thematic requirements outside the compound.

(21)



This correctly allows (17a). In so far as (17b) is concerned, *the offer* cannot get the theta role THEME if this role has already been assigned to *proposal* and, therefore, *the offer* is left without a theta role; this violation of the Theta Criterion suffices to rule out (17b).

Example (22) (where the argument is satisfied inside the compound instead of outside it), is ruled out, by our conception of g-Case, because *consider* is a Case assigner, and yet Case has not been assigned.

(22) *they are [proposal-considering]_V

Finally, in contrast with the ungrammatical (17c), (23) is grammatical.

(23) his [quick-thinking]_N saved us

This is to be expected given the lexical difference between *consider* and *think*: while the first selects a THEME and is obligatorily transitive (i.e. assigns Case obligatorily), the second does not necessarily select a TH, and accordingly does not need to be transitive (and in those circumstances is interpreted as having a generic object). Thus the contrast between (17c) and (23) is parallel to that between (24a) and (24b).

(24) a. *John considers.

b. John thinks.

To summarise, drawing on the data from English synthetic compounds, it can be observed that the notion of barrier captures the creation of a local domain within which arguments have to be satisfied.

An effect of the use of barrierhood in the example above is that it disallows the use of a zero affix to distinguish between [considering]_N and [considering]_V. If *considering* as noun were derived from the verb by affixation of \emptyset , \emptyset would not have enough features (indeed it would lack Φ -features, thus it would be featureless in the relevant sense here) for its projection to be a barrier (cf. the discussion of the *that*-trace effect). Our analysis in terms of government is inconsistent with a derivation of e.g. the pair [considering]_N/[considering]_V by zero affixation. I have argued above against some unmotivated instances of empty categories, and Lieber 1981 considers why zero affixation is a bad solution for the pairs at issue in particular. In accordance with this, the analysis adopted here (e.g. (19) and (21)) does not resort to zero affixes.

5.1.2. *Categorial constraints*

I turn now to the the second issue that Di Sciullo 1981 raises, as to whether adverbs can enter the formation of compounds; the treatment of this problem will be extended to account for other properties of compounds.

5.1.2.1. *The adverbs in -ment*

The uniqueness of the adverbs in *-ment* in Catalan (e.g. *cansadament* 'tiredly') has been noted in several of the works discussed in Chapter Two. Some of those authors consider them derivatives (cf. Cabré and Rígau 1986), while others class them as compounds (cf. Wheeler 1977, Mascaró 1986). In all the studies mentioned (with the exception of Wheeler 1977) the adverbs in *-ment* are assumed to be derived in the lexicon.

The following argument has been put forward in favour of considering them compounds: the adverbs in *-ment* present no vowel reduction, like all compounds and only some derivatives. This phonological property differentiates *cansadament* [kənsaðəmen] 'tiredly' from *cansament* [kənsəmen] 'tiredness', where the first has two unreduced vowels ([a] and [e]) and the second only one ([e]).

The Catalan adverbs in *-ment* have at their base an inflected adjective (e.g. *cansada* 'tired-FEM'), or at least a form that coincides with the [+FEM] form of the adjective. If derivation and compounding are assumed to precede inflection, we are in trouble with an inflected form at the base of a compound. Is it possible to find other cases of inflected forms at the base of a word which is unequivocally classified as a compound? Consider (25).

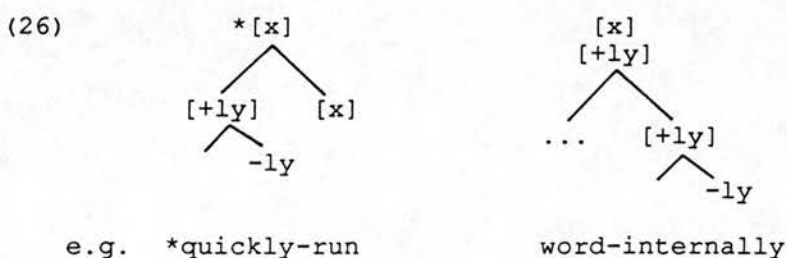
- (25) a. para-brisa
stop breeze 'window screen'
- b. salta taulells
jump counter 'shop assistant'

The evidence is, at best, dubious: in (25a) *para* can be considered basic (i.e. not inflected) just as well as inflected (i.e. third SG present indicative or second SG imperative). In (25b) the plural marker is peripheral, that is, not word-internal in the way the inflectional marker is in *cansadament*. Far from signalling them as compounds, the presence of a gender marker inside the adverbs in *-ment* suggests that they should be classified as inflectional markers; their morphological characteristics mark them as exceptional both as derivatives and as compounds.

The most conclusive argument against considering the adverbs in *-ment* to be compounds is that the element *ment* cannot be identified with the Catalan *ment* 'mind' because they do not have the same meaning. The fact that the Latin *mens*, in which it originates, means 'mind', is irrelevant here.

The factors above suggest that, for morphosyntactic purposes (although perhaps not so for phonological ones), the adverbs in *-ment* are better not classified as compounds.

The distributional singularity of the adverbs in *-ly* in English is noted already in Jackendoff 1972. To encode it, Williams (1981b) resorts to a feature [+ly]. To explain the fact that there are no compounds whose left member is an adverb in *-ly*, Williams says that, in that position, the feature [+ly] would not percolate -- but that as a head feature it must percolate and, if it cannot, the resulting word is ill-formed.



Williams' treatment of the adverbs in *-ly* is subsumed by his approach to inflection, on the assumption that *-ly* is an inflectional marker and inflectional features cannot occur inside compounds, if they do not occur in head position of the compound. At any rate, the singularity of the adverbs in *-ly* has to be independently stated (and this is done by means of a feature).

An analysis of the adverbs in *-ly* proposed by Larson (1987) will be adopted here for those in *-ment* of Catalan. The basic assumption of Larson's approach is that

nouns and adjectives must receive Case; this assumption coincides, in fact, with that of traditional grammar, where, like nouns, adjectives are Case-marked. Usually, adjectives get Case (i) by concord with a noun to which Case is directly assigned (as in *paint the red barn*) or (ii) by agreement with an NP in a predicative construction (as in *John painted the barn red*). However, in adjunct position, there is no noun from which an adjective could get Case. According to Larson, this failure of the adjective to receive Case rules out (27a), while in (27b), on the other hand, *-ly* is taken to be a Case marker which allows a "Case-dependent category" like AP to appear in adjunct position.

- (27) a. *John reasons bad.
b. John reasons badly.

There are sentences in which the position occupied by *badly* in (27b) is filled by a word not marked by *-ly*, e.g.

- (28) It rained late.
hard.
Jill drives fast.

These occur as modifiers of nouns as well:

- (29) He drives a late bus.
A hard rock fell on my head.
I have a fast car.

In (28) (and presumably in (29)), these APs are considered intrinsically Case-marked (this is expressed with the feature [+F] in Larson 1985, 1987), and so can appear in adjunct position without the presence of *-ly*. Similarly, some bare-NP adverbials are assumed to receive Case inherently; compare e.g. (30a), with (30b), where the NP adverbial is not intrinsically Case-marked.

- (30) a. John arrived that day.
b. *John arrived that period.

In Larson's words, the analysis proposed involves "the 'decomposition' of *-ly* adverbs into roughly 'prepositional' and adjectival components" (Larson 1987:253). The structure that he chooses to represent this is (31a), while (31b) represents a NP introduced by a preposition whose only function is to assign Case:

- (31) a.  b. 

I find the parallelism in (31) slightly misleading in that the role of *-ly* as established by Larson himself is more limited than that of prepositions, because even if there are prepositions whose sole function might be assigning Case (e.g. some occurrences of *de* 'of' in Catalan), it is not so for prepositions as a class. So, the AdvP in (31a) is never more than an AP marked [+Case] (or [+F]), while sometimes the structure in (30b) results in more than just the addition of Case.

Larson (1987) points out a potential problem for the syntactic derivation of the *-ly* adverbs: the lack of compositionality of some of them (e.g. *readily* *'in a ready manner'). Both in Catalan and in French the same problem arises: for instance *segurament*, from *segura* 'sure', means 'probably', not *'with certainty'. That is, if syntactically derived, these "adverbs" should be not only fully productive (as they are), but also predictable in meaning. Hence, Larson proposes to have two *-ly* morphemes in English, one syntactic, one lexical, of which only the first is affixed in the syntax. Contra Larson, I do not think that postulating two *-ly* or *-ment* morphemes achieves anything (cf. section 1.3.), since the idiosyncratic forms such as *readily* or *segurament* must in any case be listed in the lexicon as adjectives marked for [+F] (which allows their presence without another Case marker), and with their corresponding meanings. Consequently, idiosyncratic *readily*, *segurament* and the like are lexically inserted as units. To capture their formal regularity it is enough to assume a "redundancy rule" (cf. Jackendoff 1975) which can be included in the lexicon. With these two non-essential provisos, I shall adopt Larson's account of the adverbs in *-ly* in terms of Case theory.

What follows from this analysis is that adverbs do not really form a syntactic category of their own, but are rather the union of Case-marked adjectives and VP modifier PPs. In the application of this approach to Catalan, two kinds of Case-marked adjectives are found:

1. Those in *-ment*, regularly related to their [-F] counterparts, where the Case marker is phonetically overt:

(32) una tortuga lenta
 a tortoise slow
 'a slow tortoise'

Parla lentament/*lenta.
 s/he-talks slowly

2. Those which are inherently Case-marked and and irregularly related in their phonology to their [-F] counterparts³:

³I cannot find any example like those of English in (28) and (29) where an inherently Case-marked adjective is phonetically the same as a [-F] adjective. There are only signs that some adjectives are becoming [+F], e.g. *Va venir ràpidament/ràpid* 'he came quickly'.

- (33) un bon[-F]/*bé[+F] final
 a good end

 va acabar bé[+F]/*bo[-F]
 it-finished well

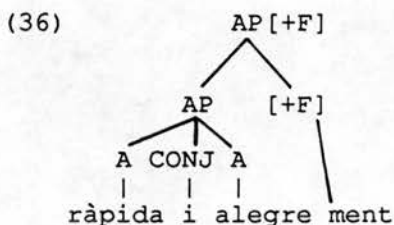
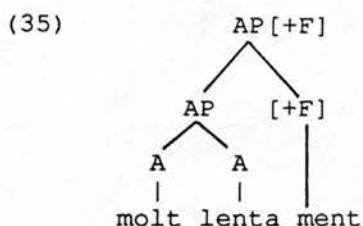
There are some other differences between adjectives in adjunct position and in adnominal position. Catalan adjectives display number and gender agreement, but only when they are [-F], modifiers as in (34a), or predicatives as in (34b). When in adjunct position, they do not display gender or number markers, as shown in (34c).

- (34) a. la tanca[+FEM] vermella[+FEM]/*vermell[-FEM]
 the fence red
 'the red fence'
- b. El Joan va pintar la tanca vermella/*vermell.
 DET Joan painted the fence red
- c. canten[+PL] bé/*béns[+PL]
 they-sing well

For compounding, this approach has the desired consequences if the assumption is made that Case markers falling within the scope of the Case Filter are not realised word-internally. Under the approach to Case theory put forward in section 4.1., it is shown that satisfying the Case Filter is not the only means of fulfilling the Visibility Condition: compounding is a particular instance of g-Case, so no compound-internal Case-marking is necessary, and thus only adjectives (i.e. non-Case-marked constituents) can take part in the formation of compounds (i.e. *[Adv V] inside an X^0 constituent), or for that matter in the formation of any word structure. . Thus e.g. *-ment* consistently does not appear internal to a compound.

Lastly, Larson argues that *-ly* subcategorises for a phrase (the same as other Case markers, for instance the genitive 's of English), and *-ment* should be identical to *-ly* in that respect. If this is so, and compounds are X^0 projections made up of other X^0 projections, not phrases, the subcategorisation requirements of *-ment* are not fulfilled inside a compound, hence once again *-ment* is not expected to appear compound-internally. This would give the structure in (35) for *Treballa molt lentament* 's/he works very slowly'. Furthermore, this syntactic structure can be readily given the right semantic interpretation (i.e. **ment** (**molt lenta**) instead of **molt** (**ment lenta**), where *molt* has scope over the adjective only, and *ment* has scope over the application of *molt* to the adjective). This would also correctly predict the standard (though not prescribed) Catalan construction *Canta ràpida i alegrement* 's/he sings quick and merrily', as in (36). As for the prescribed *Canta ràpidament i alegre* 's/he sings quickly and merry', this *-ment* is infixed to the APs that it assigns Case to⁴.

⁴On the other hand, treating the idiosyncratic *segurament* 'probably' as a unit, as proposed above, correctly predicts the oddity of ??*molt segurament* 'very probably': since *molt* 'very' does not modify *segura* 'sure', *molt segurament* does not mean 'in a very sure way'. The only possible reading is odd just in the same way as ??*I'm sure that perhaps he'll come* is.



For these reasons the compounds in *-ment* in (12c), recorded by Di Sciullo (1981), are ruled out as desired. The apparent instances of [Adv V] in French reported in (14) are in fact [A V], since *mal* is both an adjective ('bad') and an "adverb" ('badly'). In this way, the problem presented in [2] at the beginning of this chapter is given a solution without reference to the theory of government.

5.1.2.2. Extending the analysis

Some of the reasons adduced to rule out the "adverbs" in *-ment* from compound-internal positions can be used as well to rule out other ungrammatical expressions. I refer to the disallowed appearance of some "inflectional"⁵ markers inside compounds. Seen from another perspective, this kind of inflection is only found at the periphery of words (cf. Zwicky 1985).

Let us briefly recall the argument for Case markers (cf. 4.2. above). If Case assigned under the Case Filter does not apply inside X^0 structures, and a compound constitutes an X^0 projection, then a compound can only be assigned Case as a unit. Therefore, (37a) is well-formed in a language with morphological Case (where "c" stands for a Case-marking affix). By contrast, (37b) is not generated, because Case cannot percolate down to the compound-internal nominal (since the nominal is not the head), nor can the nominal be assigned Case directly by X.

⁵By this I mean what has traditionally been grouped under inflection (cf. Anderson 1982). This does not constitute a unitary phenomenon in my (and many other) analyses, and therefore has no theoretical status.

```

      [+N]
      [+Case]
      /  \
    [+N]  c
    /  \
  X    [+N]

```

$$\begin{array}{c} *X \\ \swarrow \quad \searrow \\ [+N] \quad X \\ [+Case] \\ \swarrow \quad \searrow \\ [+N] \quad c \end{array}$$

As a consequence, of all "inflectional" features, only inherent features of the compounded elements are expected to be present inside the compound. Unlike them, non-inherent features of words which are assigned to satisfy the Case Filter can be assigned to compound constituents only by (downward) percolation. If percolation is not possible, we do not expect to find those non-inherent features internal to the compound. This prediction is borne out for example in compound adjectives, as shown in (38), where the relevant features are gender and number in adjectives.

- Adjectives agree in gender and number with the nouns they modify, in Catalan, but, no matter what the gender or number of the modified noun, the first member of the compound appears invariably in the unmarked form, that is, as [-FEM] and

- (42) a. *sang-er
sing-PAST -er
'somebody who sang/used to sing'
- b. *parava-llamps
s/he-stop-PAST lightning
- *trencarà-nous
s/he-break-FUT nuts

The facts of compounding become quite natural in a framework (such as GB) in which INFL is treated as an independent constituent. To recapitulate, then, the absence of verbal inflection in compound-internal position is accounted for differently than the absence of other morphosyntactic marks, such as agreement markers. However I see no way of extending these approaches to other "inflectional" phenomena, such as inherent morphological markers. For instance, I have ruled out compound-internal gender and number markers when they are the result of agreement, but the account above cannot rule out gender and number when they are inherent properties of a category. So, nouns are [+FEM] or [-FEM] without that being a matter of agreement, and therefore the feature [+/-FEM] is not forbidden compound-internally. Compare (43) to (38b).

- (43) cama-trencar
leg[+FEM] break-INF 'to break some-one's/
thing's leg/s'

*nassos-pelats
 nose-PL peeled-PL 'with peeled noses'

What remains to be explained, then, is why the expressions in (44c) are ungrammatical⁶. The argument must be that the cause of the ill-formedness of (44c) is not going to be the same as the one for (37b), (38b) and (42), which have already been accounted for on the basis of Case theory and the status of INFL. To my knowledge, examples like (37b), (38b) and (42) are universally ungrammatical, while whatever rules out (44c) does not seem to hold universally, since grammatical parallel examples have been found, e.g., in English, as shown in (45) (taken by Borer from Kiparsky 1982).

- (45) a heads-up play
 a hands-off policy
 the excess profits tax
 lice-infested

5.1.2.4. Some other issues of syntactic category

Even though it has been ruled out to have an inflected verb inside a compound, what rules out a structure such as the one in (46)?

- (46) *[INFL X]_X

In (46) INFL does not appear as inflection on a verb, but rather as a base-generated constituent of the compound. Like (46), the compound structures in (47) give rise to ungrammaticality.

- (47) a. *[DET X]_X
 *[CONJ X]_X
 *[COMP X]_X
 b. *[X DET]_{DET}
 *[X CONJ]_{CONJ}
 *[X COMP]_{COMP}

⁶Suppose that [+PL] is handled in a fashion parallel to INFL -- i.e. assuming that number is an independent syntactic projection, as proposed by Szabolcsi 1987, etc.; then the ungrammaticality of (44b) would be accounted for, but the grammaticality of (44a) would become problematic.

I shall argue that all the structures in (46) and (47) are out for the same reason⁷. It seems that only words defined for the features [+/-N] and [+/-V] enter the composition of compounds. These are the major categories N, A, V, and P. In (46) and (47) constituents are included which are not defined for these two features: INFL, DET, COMP, CONJ. Short of an explanation for this, it seems necessary to stipulate that compounds have constituents belonging to the major categories N, A, V and P (i.e. elements defined for the features [+/-N] and [+/-V]).

Categories marked positively for the features [+/-N] and [+/-V] are often referred to as "lexical categories"; then, a preposition is not lexical in the sense that it is negatively characterised for both features: [-N] and [-V] -- cf. Chomsky 1981⁸. Like prepositions, notice that CONJ, DET, INFL and COMP all belong to closed classes of words, which are not freely generated -- i.e. they are included in a grammar's lexicon, but not freely produced by the grammar itself. Hence, no compound COMPs, CONJs or DETs are expected, the same way that there are no COMPs, CONJs or DETs formed by affixation. It is predictable, from this perspective, that the structures in (47b) are impossible as compounds. On these grounds, [X P]_P can also be ruled out: prepositions belong to a closed class of words, and so no new prepositions are formed by composition. However, prepositions take part in compounding (in compounds of the *sotabarba* type, for instance).

An interesting correlation between these categorial restrictions and their semantics can be observed: nouns, adjectives and verbs, when intransitive, are the categories that denote properties; precisely these are the categories that compounding results in. Arguably, only the expression of categories denoting new properties can be envisaged within a given grammar.

To recapitulate, there are two classifications of words which are both necessary to determine the possibility of compounding: (i) major categories (determined by [+/-N] and [+/-V]) vs. minor categories, and (ii) lexical categories (N, A, V) vs. non-lexical categories. Only words which belong to a major category can compound (in a technical sense). Only lexical categories can be freely generated (i.e. they can be identified with open-class words). Because prepositions belong to a major category but are not lexical, they are not freely formed, but can be constituents of a compound (i.e. they cannot be heads in compounds, but they can enter compound formation).

⁷In the circumstance of INFL being an affix, (46) is ruled out for the extra reason that its subcategorisation requirements are not satisfied.

⁸A different use of the term "lexical" is made in e.g. Chomsky 1986b. There, prepositions L-mark their complements, and L-marking is theta-government by a *lexical* category. Hence, in *Barriers*, prepositions are included among the lexical categories. Because this is a purely terminological conflict, it has no consequences so long as what is meant by "lexical" is made clear every time the term occurs.

If we take these categorial properties of compounds seriously, the expressions in (48) cannot be included among compounds, since they include CONJ and DET:

- (48) a. *puja-i-baixa*
 go-up and go-down 'action of going up and down'
- plats-i-olles*
 plates and pots 'plates and pots shop'
- b. *a-les-hores*
 in the hours 'then'

It is beyond the scope of this dissertation to classify the French expressions with DETs, such as *trompe-l'oeil*, literally 'deceive-the-eye', 'illusion'. As for the Catalan well-formed expressions in (48), they cannot be considered X^0 projection expressions, though they are lexicalised and thus interpreted very much the same way as X^0 projections. As noted in Zwicky 1985, having encountered lexicalised forms for all other major categories, it is not surprising to find them for nouns of X' and XP projection too.

As predicted, then, the lexicalised expressions exemplified in (48) present the same characteristics as other non-lexicalised expressions of the same category and bar-level. In particular, the requirements for those in (48a) are those which apply to coordinated structures -- minimally, same syntactic category. Because X'/XP expressions are freely generated, new lexicalised X' expressions appear in the language due to use, and acquire specific meanings which are not entirely predictable from the linguistic form alone. I shall come back to lexicalisation in Chapter Six.

In the preceding subsections I have addressed question [2] formulated at the beginning of the chapter (namely, are all or some Adv ever constituents of compounds?), which arose in the context of Di Sciullo 1981. The ill-formed compounds in (12a) and (12c) have been ruled out on the basis of factors independent from government.

The $*[NP\ V]$ compounds of (12b) could be ruled out if the No Phrase Constraint were adopted, though Di Sciullo claims that it does not hold for French. Moreover, the point that she makes with examples such as *l'oeil-trompe* is that external arguments do not appear as complements in compounds. Baker (1988a) accounts for facts similar to these by analysing them as instances of incorporation. Incorporation of an external argument is impossible because the trace of the incorporated nominal is not properly governed (and therefore the ECP is violated). This correctly predicts the non-existence of incorporation of external arguments in quite a number of languages. However, if movement does not occur in French and Catalan compounding (as I claim), this transformational account is not available to rule out (12b). Clearly, what is needed is some reason to filter out external arguments as complements in compounds.

- (49) *[[+N, -V] [X]]_x
 where [+N, -V] corresponds to the external
 argument in X's θ -grid.

From a non-transformational perspective, one can argue that the distinction between external and internal argument is precisely that an external argument is assigned by the projection of the theta role assigner, i.e. higher in the tree and not as sister of the theta role assigner. In that case, the [+N, -V] constituent in (49) cannot correspond to the external argument of X, being as it is the sister of X. Recall that the lexical entry of a theta role assigner includes a theta-grid where the external argument is marked as different from the internal arguments. If this is correct, then the ungrammaticality of (12b), where NP satisfies the external argument, is motivated. Just as with (12a) and (12c), though for different reasons, this account does not rest on the theory of government.

5.2. Universal constraints in morphology revisited

Having looked at the way in which several subtheories of a GB grammar (theta theory, Case theory, binding theory, X-bar theory and government theory) can interact to predict word formation phenomena, we are in a position to reformulate some principles of morphology. (These were summarised, after Shibatani and Kageyama 1988, in section 3.2. in Chapter Three).

1. The fact that Case particles can be absent from nouns which are part of a compound is entirely predicted by the approach to Case theory in section 3.1., following Baker 1988a. Case particles are (language-specific) manifestations of g-Case regulated by the Case Filter. However, the Case Filter does not apply to nouns inside compounds, precisely because compounding is an alternative means of identification (cf. the Visibility Condition). Therefore, compounds fall within the scope of g-Case, but not within the scope of the Case Filter. In actual fact, under the approach here, Case particles must be absent from nouns once they are part of a compound. I am not aware of any piece of data contrary to this.
2. That tense inflection is banned from the components of a compound is accounted for, in a GB grammar, by the fact that I (or INFL) appears as a syntactically independent node in a sentence structure -- indeed as the head of IP, i.e. a sentence without a complementiser. Tense (like any verbal inflectional features, mood, etc.) is a sentential feature, and its presence implies the existence of the whole sentence,

not only an X^0 projection. Briefly, INFL is not a candidate as element for the formation of compounds.

3. X^0 constituents resist syntactic interruption. "Base-generated" intervening constituents are disallowed for purely structural reasons: they would be sister constituents of a head, and interpreted as such. As for the application of Move α , this framework in its present formulation does not allow for movement into X^0 projections without giving rise to ECP violations.
4. That the only structures postulated in our theory of word formation are binary is now presupposed by the feature percolation conventions adopted (see section 4.4.2.). No way has been found of deriving binarism in word formation from what has been said here; it is possible to find, however, some empirical evidence for the adequacy of binarism, as pointed out by Pesetsky (1985): "Were [binarism] not true, we would expect to find affixes that demand, as a positional or categorial requirement of subcategorization, two sisters" (Pesetsky 1985:215 note 17). For example, $A + B + \text{affix}$ would be grammatical, where A and B are morphemes, but not $A + \text{affix}$, or $B + \text{affix}$.
5. When compared to the various renderings of combination of an argument-taking category with an argument (cf. Roeper and Siegel's First Sister Principle (1978), Selkirk's First Order Projection Condition (1982), etc.), the perspective taken here has been quite different. It has been assumed that the PrPr holds for compounds and that, therefore, the internal arguments of their argument-taking constituents have to be satisfied. Some considerations with respect to realisation of theta roles (e.g. the need that all internal theta roles be realised at the same level in a tree) may need to be expressed, but these principles hold for compounds and phrases equally. Finally, a few language specific considerations are necessary to account for the differences between, say, Catalan and English, with respect to compounding.
6. As already pointed out, word formation is idiosyncratic only insofar as the lexicon is. On the other hand, the application of syntactic rules and principles is not subject to statements of exception. Putting the two together, it follows that word formation, if taking place in the syntax, is not "lexically governed".

Next I comment briefly on the three properties that, according to Shibatani and Kageyama (1988), mark the contrast between lexical and postlexical (i.e. syntactic and postsyntactic) compounds.

7. Only postlexical compounding can involve maximal projections, as in the structure in (50a). (It is understood that the head of a compound must be an X^0 projection). This licenses the principles of X-bar syntax to apply fully within the X^0 level. Lexical compounds, on the other hand, do not allow phrases among their constituents, as in (50b). Catalan compounding, which I claim to be syntactic, also disallows phrases as constituents of a compound.

- (50) a. $[XP\ Y]_Y$
 b. $[X\ Y]_Y$
 $*[XP\ Y]_Y$
 (order irrelevant)

Therefore, given the variation in the data concerning this phenomenon, it seems advisable to keep the No Phrase Constraint as a language specific filter -- applying, for example, to Catalan. I shall come back to this issue in Chapter Seven.

8. Shibatani and Kageyama (1988) claim that anaphoric relationships can hold between a constituent of a postlexical compound and an element outside it. This has also been my claim for Catalan compounds. Lieber (1984) points to the possibility that the binding theory should be allowed to apply even inside complex words formed in the lexicon, once these appear in a syntactic tree. She gives compelling evidence that there is no reason to stipulate anything in this respect: the theory of binding itself suffices to rule out unwanted anaphoric bindings. Thus no stipulation is required to get the desirable predictions.
9. It has been found true for Catalan that the constituents of a complex word do not have a referential function, but rather a generic one. But if this is not so for postlexical Japanese compounds (cf. Shibatani and Kageyama 1988), this property cannot be taken as definitional for words. Nor can it be derived: Sproat (1985) holds that only maximal projections can refer, not X^0 projections; however, if the data on Japanese has been rightly analysed, the solution offered by Sproat cannot be the correct one.

To recapitulate: of all the characteristics of morphological forms stated above, quite a number ([1], [2], [3]) follow from the independently motivated principles of GB (e.g. of Case and theta theory). Some others ([7] and perhaps [9]) do not hold

universally and, therefore, we do not want to make them an inherent part of a theory of morphology. Finally, another ([6]) has been rejected in view of what has been shown previously. An account in terms of GB replaces previous ones in the case of [5] and [8], and [4] is taken as a premise of the formalism.

I shall come back to morphological constraints as part of morphological theory in Chapter Seven.

Chapter Six

Catalan compounds

This chapter contains a comprehensive analysis of complex words in Catalan within the general framework set up in Chapters Three, Four and Five. I address the data considered in Chapter Two, though not all the expressions there will be taken to be compounds proper here. For those which are not, alternative structure assignments will be suggested; in this way, a principled classification of Catalan complex expressions is offered.

6.1. The right-headed compounds of Catalan

In their recent review of word formation, Di Sciullo and Williams write: "It now appears that French (and no doubt Spanish) lacks compounding altogether. Once we have subtracted fixed syntactic phrases (idioms) such as *timbres-poste* and phrases reanalysed as words (syntactic words) such as *essui-glace* (sic), there are no candidates left" (Di Sciullo and Williams 1987:83)¹. In this chapter I shall show that this cannot be maintained for Catalan, where some constructions are best considered compounds. On the other hand, Di Sciullo and Williams' move, to reclassify constructions formerly considered compounds, is made here, too. Although it is not my goal to compare Catalan with its sister Romance languages in any detail, I shall include here a few suggestions as to why a contrast arises between Catalan on the one hand and French and Spanish on the other, whether or not it is true or not that these last two lack compounding.

We will begin by reviewing the properties of compounds as exemplified by

¹Di Sciullo and Williams' syntactic phrases correspond to the lexicalised expressions examined in 6.4. Syntactic words are compounds of the *parallamps* type, and these authors' analysis is reported in 3.3.2.2.

camatrencar 'to leg-break'. Since neither of its constituents is an affix, *camatrencar* is not a derivative. As shown in (1), the compound exhibits inflectional variation peripherally, but not internally, in contrast with the constituents of phrases.

- (1) *cama trenc ar*
 leg break INF 'to leg break'
- cama trenc o*
 1-SG 'I leg-break'
- cama trenqu es*
 2-SG 'you leg-break'
- **comes trenc ar*
 leg-PL break INF

Syntactically, it interacts with other constituents as word units do, so that interruption or modification of its constituents is disallowed, as seen in (2) and (3).

- (2) **Es cama-massa-trenca.*
 SE leg too-much breaks
 'S/he leg-breaks too much'
- (3) **Es cama-curta-trenca*
 SE leg short breaks
 'S/he leg-breaks a short one'
- **Es curta-cama-trenca*
 SE short leg breaks
 (same meaning)

The analysis of Di Sciullo and Williams (1987) reclassifying a putative compound as a "syntactic word" or as an idiomatic syntactic phrase is inadequate here. These alternatives are only at hand if there are phrases in the language under inspection with precisely the same form as the "words" at issue. That is, *essuie-glace* is considered a phrase reanalysed as a word because there are phrases in French with the same characteristics (e.g. *Il essuie les glaces* 'he wipes the windows'), except that they are not lexicalised. And the same goes for idiomatic syntactic phrases. In the case of *camatrencar* these alternative analyses have to be excluded because there are no phrases with the shape of *camatrencar*, where a verb is preceded by its direct object (e.g. **Un vidre he trencat* 'a window I have broken' vs. *He trencat un vidre* 'I have broken a window'). Hence, *camatrencar* cannot be a lexicalised phrase.

Consequently, the grammar of Catalan must encompass some mechanism for the generation of compounds, i.e. minimally $[X^0, Y^0]_{Y^0}$ structures. If we find an instance of compounding, e.g. *camatrencar*, the next question is: Are the mechanisms that generate it limited to producing this kind of compound? Or are these mechanisms capable of generating other kinds of compounds?

I shall suppose, following the lines of Lieber (1983) among others, that

concatenation applies freely and of itself overgenerates, with a set of principles defined that rule out ill-formed complex words. This is in consonance with the Principles and Parameters framework adopted here. I shall systematically examine all possible combinations of X^0 categories and exclude those which do not occur in Catalan on the basis of the principles made explicit in GB. Ideally, we should account both for all possible compounds and the gaps in the lexicon on the basis of those principles.

Camatrençar is right-headed, and my point of departure will be the assumption that Catalan compounds are consistently right-headed. Apparent counterexamples will be discussed in section 6.4. In section 5.1.1.4. I have indicated that the categories that enter compound formation are the four major categories. Their combination gives 16 possible right-headed compound structures:

(4)

	N	V	A	P	(second)
N	[N N]	[N V]	[N A]	[N P]	
V	[V N]	[V V]	[V A]	[V P]	
A	[A N]	[A V]	[A A]	[A P]	
P	[P N]	[P V]	[P A]	[P P]	
	(first)				

There are two semantic relations that the theory allows for in these configurations: (i) the relation between an argument-taking category and its argument, and (ii) the relation between a modifier and a modified category. The head is the argument-taking category in the first case, and the modified category in the second.

6.1.1. [N N]

Some existing words which are candidates for the $[N N]_N$ structure are listed in (5).

(5)	figa-flor	
	fig flower	'weak character'
	ferro-carril	
	iron rail	'railway'
	lliri-jonc	
	lily rush	'sword-lily'
	vagó restaurant	
	wagon restaurant	'restaurant car'

paper moneda	
paper coin	'bank note'
peix martell	
fish hammer	'hammer fish'
pruna clàudia	
plum Claudia	'greengage'
targeta postal	
card postcard	'postcard' ²

Given that most nouns do not obligatorily assign a theta role to any complement, the noun preceding the head can only be expected to be a modifier. As observed by Dowty (1979), cf. section 3.3.1., the precise interpretation of the compound cannot be narrowed down in any way other than pragmatic. Thus the words in (5) take on interpretations in a broad and unpredictable fashion. This is also the case with compounds of this kind in English, e.g. *fire truck*, *desk stamp* and so on (cf. Lieber 1983:255).

Looking in some detail at the examples in (5), discrepancies can be found in their morphosyntactic properties, which become apparent in the plural counterparts given in (6).

- (6) a. figa-flors
fig flower-PL
- *figues-flors
fig-PL flowers-PL
- ferro-carrils
iron rail-PL
- *ferros-carrils
iron PL rail-PL
- b. *paper monedes
paper coin-PL
- papers moneda
paper-PL coin
- *peix martells
fish hammer-PL
- peixos martell
fish-PL hammer

²I categorise *postal* as a noun, here, though in some contexts it is an adjective, e.g. *oficina postal* 'postal office'.

c.*targeta postals
 card postcard-PL

targetes postals
 card-PL postcard-PL

Because it is required of the grammar that it have predictive power for the morphosyntactic characteristics of complex words, the words in (6a), (6b) and (6c) cannot belong together.

I consider compounds to be morphosyntactic units, so that they will have peripheral (non-inherent) plural markers, affixed to the right of the word (in Catalan), as in (6a). The expression in (6c) pluralises just like a phrase, and in fact e.g. *targetes* 'cards' has often been taken to be the head, so that *targetes postals* can be straightforwardly considered a lexicalised X'. The same holds for *papers moneda*, parallel to the French *timbres poste*, which Di Sciullo and Williams (1987) also analysed as a lexicalised "phrase". The difference between (6b) and (6c) is that in (6b) only the first element, the head, is pluralised. This kind of example is quite rare and its occurrence is unpredictable. However, the fact that the second element is invariable indicates, once more, that the expression is lexicalised. Corroborating the idiosyncrasy of (6b) there is the example *pruna clàudia* 'greengage', whose plural is *prunes clàudia*, but more often *prunes clàudies*, according to Badia (1962). This indicates that the group illustrated in (6b) is unstable and that its members are likely to be reclassified by speakers as belonging to the class of (6c). The unmarked cases remain (6a) and (6c), the first compounds strictly speaking, the second lexicalised X' expressions. The morphosyntactic behaviour of both is entirely predictable from the structure that they have been assigned.

6.1.2. [V N]

There are no righthanded compounds of this form and new ones cannot be coined:

- (7) *seu-cadira
 sit chair '?chair to sit on'
- *menja-pa
 eat bread '?bread ready to be eaten'
- *perdona-persona
 forgive person '?person who forgives (easily)'

Theta theory suffices to rule out most [V N]_N compounds, because the verb is in non-head position, where it cannot have its arguments satisfied. The verbs in (7) have different theta-grids, but they are all ungrammatical for the same reason, namely that the position of the verb is such that no arguments can be realised, in violation of the Theta Criterion, and the PrPr.

Yet, for a small number of verbs, the problem remains. Suppose we have a compound with the structure $[V N]_N$ whose verb has no theta role to assign, e.g. *ploure*, 'to rain'. Then one must appeal to principles other than the Theta Criterion to rule out, e.g., $*[plou\ dia]_N$ 'rain-day'. What would be the interpretation of the compound? The verb must be either an argument of the head noun, or its modifier. Following Kayne (1982), I assume that there is complementarity of properties of nouns and verbs, and that bare verbs cannot be arguments nor modifiers. As a consequence, $*[plou\ dia]_N$ and the like can be ruled out because *plou* cannot be a modifier or argument of the noun.

6.1.3. $[A N]$

I have been unable to find many existing compounds of this type. (8) groups expressions which are different in various respects.

- (8) a. *mal-son*
bad sleep 'nightmare'
- sant crist*
holy christ 'crucifix'
- b. *poca-pena*
little-FEM 'sorrow' 'miserable person'
- poca-solta*
little-FEM reason 'thoughtless'
- c. *mala-lluna*
bad-FEM moon 'bad mood'
- *mal-lluna*
bad moon
- mitja-nit*
half-FEM night 'mid-night'
- *mig-nit*
half night
- d. *lliure-canvista*
free trade-ist 'free-trader'
- bon-homia*
good man-hood 'good nature'

(8a) contains compounds of this type *stricto sensu*. These are simply analysed as instances of modification, i.e. the adjective modifies the head noun. (8b) correspond to "exocentric" expressions (cf. 3.3.2.2.), and (8c) to lexicalised X' (one of the diagnostics to determine their category is the presence of agreement markers, such as [+FEM] on the adjectives of both examples). The expressions in (8d) are derived

by suffixation from *lliure canvi* 'free trade' and *bon home* 'good man'³.

The lexicalised *X'* in (8c) contrast with the ones encountered in (6c), in their plural counterparts:

- (9) mitja-nits
half night-PL

?mitges-nits
half-PL night-PL

I shall come back to lexicalised expressions in 6.4.

Although there are relatively few compounds of this type, it seems to me that the grammar of Catalan provides one the possibility of creating new compounds like these, but does not entail the necessity of really creating them; this necessity cannot be accounted for in purely linguistic terms.

6.1.4. [P N]

Leaving aside the "exocentric" compounds of the type of *sotabarba*, 'under-chin', 'double chin' (see section 3.3.2.2.), there are no compounds of the form [P N]. This is predicted by theta theory, because a preposition in non-head position cannot satisfy its arguments⁴.

- (10) *sota-salt
under jump '?jump which is under something'
- *entre-primavera
between spring '?spring in between'

Consider next the words in (11):

- (11) en-bagular
trunk-INF 'to put in a trunk'
 or 'to fill with trunks'

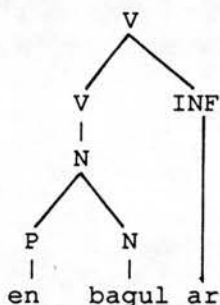
³Notice that this implies that, in my analysis, e.g. *lliure-canvista* 'free-trader' has a structure which can be represented as [[lliure canvi] -sta], not [[lliure [canvista]]]. This bracketing does not respect word boundaries, but does not give rise to bracketing paradoxes either, since the morphosyntactic structure is the same as the one indicating semantic scope. (This is an instance of syntactic affixation, cf. Fabb 1984).

⁴The fact that similar examples (e.g. *in-step*, *out-side*) are good in English only reflects on the different properties of English prepositions, which unlike their Catalan counterparts, can be intransitive.

a-pilar		
pile-INF		'to make a pile with'
en-coloniar		
cologne-water-INF		'to put cologne-water on'
en-sucrar		
sugar-INF		'to put sugar on'

There are two possible analyses for (11): one perhaps as compounds, the other as derivatives. If the words in (11) are compounds, they are assigned the following structure:

(12)



This analysis runs into problems in so far as it involves not only compounding, but also category change, unlike the other instances of compounding. If *embagular* has a [P N] compound at its base, why do we never encounter *embagul* as a N? Besides, as argued above, theta theoretic principles rule it out: *en* would not assign its theta role in non-head position. (That *en* assigns a theta role is shown by the ungrammaticality of e.g. **Vaig anar en* 'I went in').

Together with these syntactic reasons to exclude (12), there is a phonological one: the prepositions which supposedly appear in the words in (11), *a* and *en* 'in', are for the most simple phonological variants of the same preposition (together with *amb*). The preposition *a* introduces an N' starting with a consonant, and *en/amb* one starting with a vowel (cf. Ferrater 1981a).

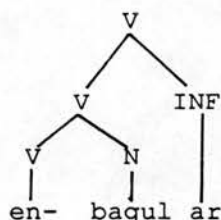
(13) Sóc a casa.
I-am at home

El vaig veure en una pel·lícula.
him I-saw in a film
'I saw him in a film'

Yet, in the words in (11) the alternation of *a*, *en* does not follow this pattern, suggesting that they are derivational affixes rather than prepositions.

Alternatively, the words in (11) are given the structure in (14).

(14)



Here the resulting word gets its category by simple percolation (the prefix is the head, following Convention II of Lieber (1980, 1983) and, because it is of category V, this is the resulting category of the word). This analysis does not run into the problems of the previous one. The only drawback is that, because *a-*, *en-* are not analysed any more as the preposition *a/en* 'in', the meaning 'in' which the paraphrase of the complex word involves must be included again. So *a-/en-* as prefixes and *a/en* as prepositions are related to the meaning 'in' separately. I would like to argue that their semantic relatedness is truly accidental synchronically, though not in the history of Romance; that is, as far as the intuitions of native speakers go, *en-* in *embagular* and *en* as preposition are no longer related. Therefore, the diachronic relation between $[en]_p$ and *en-* is something that a GB grammar does not express.

6.1.5. $[N V]$

My rendering of the *alatrencar* type in section 3.3.2.1. makes the inalienability of the first element of this compound predictable. Here are some other instances of this type.

- (15)
- | | | |
|----------------|--|--------------------------------------|
| cor-prendre | | |
| heart take-INF | | 'to captivate' |
| cor-secar | | |
| heart dry-INF | | 'to dry out somebody's life, energy' |
| coll-trencar | | |
| neck break-INF | | 'to break somebody's neck' |
| coll-vinclar | | |
| neck bend-INF | | 'to bend somebody's neck' |
| cap-ficar | | |
| head put-INF | | 'to worry' |

In Chapter Two a couple of examples are found which are problematic for the present account: *peucalcigar*, 'foot-step on', under the interpretation 'to step with one's foot', and *collportar*, 'neck-carry', under the interpretation 'to carry on one's neck'. In my judgement the former is out, indeed *calcigar* is not in my vocabulary, and the latter only has the pragmatically odd reading 'to carry somebody's neck'. The data here can be found in Mascaró 1986 and Cabré and Rigau 1986.

- (16) El cavall va peu-calcigar en Pere.
 the horse foot-stepped DET Pere
 'The horse stepped on Pere'

In (16) *peu* is not the THEME of *calcigar*, but rather seems to correspond to some semantic theta role, possibly INSTRUMENTAL, 'with one's foot'.

The second problematic example relates to the first: it is *collportar*, where *coll* is not the THEME₁ of *portar*, as shown in (17). (The cognate French word *colporter* is presumably opaque, not perceived as related to *cou* and *porter*).

- (17) El pare va coll-portar la nena fins a casa.
 the father neck-carried the girl to home
 'The father carried the girl home on his shoulders'

For this case, one might equally think that *coll* is bearing a theta role (LOCATIVE or MANNER, for instance) or not. Because all other instances of [N V]_V compounds involve compound-internal THEMES, and not any other theta role, I propose that these two examples do not involve compound-internal theta role assignment, but rather are perceived as simple words. Otherwise, we would have to account for the ungrammaticality of (18) without being able to maintain that only THEMES appear in initial position of the verbal compound.

- (18) *rem-navegar
 oar navigate-INF 'to row'
 *avió-arribar
 plane arrive-INF 'to arrive by plane'

So, again, I consider these two compounds to be exceptional and assigned a meaning in virtue of belonging to the lexicon, but without conforming to the pattern of compound formation in Catalan. Its first position is then not a theta position. (This is very much the analysis that Pesetsky (1985) proposes for the English *homemade*).

Returning to the regular kinds, although the relation of inalienability of the first constituent is predictable (cf. section 3.3.2.1.), I want to concentrate now on the fact that this constituent satisfies the theta role in the verb's theta-grid, in particular the THEME₁, and not another theta role.

Pesetsky (1985) considers the satisfaction of arguments inside words, in relation to some compounds of English involving theta role assignment. He points out that theta-marking is contingent on the existence of a structural position that must or can be theta-marked. From this, and given data of the kind of *Pasta-eating in trees is discouraged*, he concludes that the first member of a compound is a theta position. This also holds for Catalan.

The point where my approach diverges from Pesetsky's is related to X-bar theoretic assumptions. He writes: "When an element receives a theta role from V, it is normally a sister of a projection of V" (Pesetsky 1985:235)⁵. Here I assume that in a compound a verb can satisfy its argument in its sister, and not necessarily in a sister of its projection. Thus the X-bar structure of compounds is "flattened" when compared with phrasal X-bar structure. Theta role assignment among sisters is far from unprecedented; Baker (1989) summarises "the standard conditions on theta role assignment (adapted from Chomsky (1986a, 13-14))":

- (19) α may theta-mark β only if
 a. α and β are structural sisters, or
 b. a projection of α is a structural sister of β .

To capture the constraints on theta role assignment, my account of compounding in Catalan includes a well-formedness condition thus:

- (20) In the configuration $[N\ V]_V$, N can only satisfy
 a TH_1
 (in Catalan).

(20) correctly rules out (21), where other theta roles are being realised.

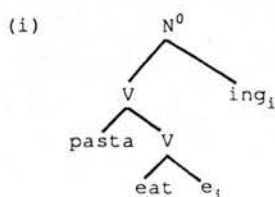
- (21) *gos-olorar
 dog smell-INF
 (EXP)

 *noi-dormir
 boy sleep-INF
 (AG)

 *platja-anar
 beach go-INF
 (GOAL)

Yet, one might argue that (20) is arbitrary in that it suggests that it is possible for a natural language to have a well-formedness condition like (20) except in that it makes reference to AGENT, EXPERIENCER, LOCATIVE, GOAL (or any other theta role) instead of $THEME_1$. In other words, using a device such as (20), even if descriptively adequate for Catalan, has implications for universal grammar that are, perhaps, objectionable. (For reference to various views on this matter, see Newmeyer 1986:206).

⁵Thus his analysis of this kind of compound involves the introduction of a structure such as that (i).



First, it must be pointed out that a cross-linguistic study would be necessary to prove that statements of the kind of (20) making reference to other theta roles are not found. In fact, a number of them are mentioned in the literature. For example, in Southern Tiwa THEMES incorporate while GOALS do not (all other factors being equal) (Baker 1988:453); Mithun reports that most languages with incorporation allow only incorporation of patients (or THEMES); others, such as Nahuatl, Comanche and Takelma, allow LOCATIVES and/or INSTRUMENTALS as well (Mithun 1984:875).

Second, it would be possible to embed the present study in a different theory of thematic roles, and modify (20) accordingly. For instance, (20) would appear less ad hoc if it did not appeal to THEMES, but to the most "passive" of all roles in a continuum of thematic features. Dowty takes up the linguistic tradition engaged in characterising prototypical roles, and in Dowty 1989 proposes a theory of thematic roles which does not include discrete thematic roles, but rather a gradual classification of roles depending on a series of semantic properties.

Third, within GB, some notions of Case theory imply that in fact conditions like (20) are found in the grammar. The idea of semantic Case and inherent Case implies that there are principles directly linking cases and particular theta roles. Then Catalan verbal compounds are nothing but another instance of semantic Case. It is generally accepted that even languages without semantic Case require some statements linking theta roles and cases (these be along the lines of Carrier-Duncan (1985), or the CSRs of Chomsky (1981)).

Reference to particular theta roles in lexical rules has been rejected by Levin and Rappaport (1986)⁶. When studying adjectival passives, they write that "operations on lexical representations may not have access to theta role labels but may only affect the process of theta role assignment" (Levin and Rappaport:657). (Later work, e.g. Borer 1990, also leads to this result). They admit that reference to theta role labels is needed for theta role assignment, inherent Case, and other syntactic phenomena. Insofar as I propose that compounding takes place in the syntax and not in the lexicon, my analysis does not conflict with Levin and Rappaport's. In fact, if the point they make is correct, having compounding in the syntax becomes not only preferable, but an absolute necessity, since only in the syntax are theta role labels available.

⁶Rappaport et al. (1987) and Rappaport and Levin (1988) come to the same conclusion, but this is not surprising given that they dispense with theta roles tout court. For this reason I shall consider Levin and Rappaport 1986 only.

6.1.6. [V V]

As in the case of [V N]_N, the verb which is the first member of this compound cannot have its arguments satisfied, being as it is in non-head position. Also, Kayne's (1982) consideration that verbs cannot be arguments or modifiers rules out a [V N] compound with a non-argument-taking verb. Hence the ungrammaticality of the examples in (22):

- (22) *salta-venir
 jump go-INF '?'to go in a jump'
 *escalfa-cuinar
 heat cook-INF '?'to cook with heat'
 *enreda-dir
 lie say-INF '?'to lie'

(However, for compounds of this form which involve coordination, see section 6.2.).

6.1.7. [A V]

Several compounds of this kind have been encountered, such as:

- (23) mal-parlar
 bad speak-INF 'to speak ill of'
 car-vendre
 expensive sell-INF 'to sell for a lot of money'
 mal-tractar
 bad treat-INF 'to ill-treat'
 prim-filar
 thin spin-INF 'to split hairs'
 mal-menar
 bad lead-INF 'to damage'
 mal-fiar-se
 bad rely-on-INF REFL 'to distrust'
 mal-avesar
 bad accustom-INF 'to (cause to) acquire
 a bad habit, to spoil'
 mal-criar
 bad bring-up-INF 'to spoil'
 mal-baratar
 bad exchange-INF 'to squander'

mal-casar
bad marry-INF

'to marry badly'

Semantically, most of these compounds are transparent. They consist of an adjective modifying a verb, and are correctly predicted to be grammatical. Notice that the structures postulated involve non-Case-marked adjectives (cf. 5.1.1.1.). That Case-marked adjectives are not allowed is shown in (24):

- (24) *malament-parlar
 badly speak-INF
- *carament-vendre
 expensively sell-INF

That *mal* is not inherently Case-marked is illustrated in (25):

- (25) Parla malament [+F]/*mal [-F]
 S/he speaks badly/*bad

A problem arises with the words in (26), where the verb is preceded by a form which looks like an adverb, i.e. a Case-marked adjective, as illustrated in (27), despite the fact that nominals do not need Case in a compound. *Bé* and *ben* are positional variants of each other.

- (26) ben-parlat
 well spoken
- ben-vingut
 well come
- ben-aventurat
 well ventured 'blessed'
- (27) Parla bé [+F]/*bo [-F]
 speaks well/*good

Ben- is the only inherently Case marked adjective which appears in this position; its [-F] counterpart, *bo*, is never the first member of a compound:

- (28) *bo-parlat
 good spoken
- *bo-vingut
 good come
- *bo-aventurat
 good ventured

Two questions arise. Firstly, why are the examples in (26) good, and secondly, why are the examples in (28) bad.

In relation to the first question, it may be noted that the words in (26) only exist

as participles, and not in the other forms of the verbal paradigm: ?*benparlar*, ?*benvenir*, ?*benaventurar*. This had led some linguists to reclassify *ben* as an affix: then the restrictions on its appearance are encoded in the affix's subcategorisation frame. Indeed, the obligatory presence of the participial form gives us a clue as to the possible origin of the words in (26): an adjectival phrase, where the adjective is deverbal, and is modified by an "adverb", as in (29).

- (29) va [molt [ben [vestit]]]
 goes very well dressed
 'He's very well dressed'

The expressions in (26) are idiosyncratic in meaning (for instance, *benaventurat*, literally 'well-ventured', means 'blessed'); they originate in phrases and have become part of the lexicon. It is possible to maintain that they are compounds in the loose sense of the word (i.e. having at their base free words, not affixes), and not derivatives. Technically, I shall consider them lexicalised structures (see section 6.4.).

Given that (30) is grammatical and parallel to (29), could *malparlar* and the like be analysed as lexicalised forms the same as *benparlat*?

- (30) Va ser mal-tractat.
 he-was bad treated
 'He was badly treated'

I want to argue against this line of thought for the following reasons: (i) *Malparlar* and the like are transparent and thus unlikely to be lexicalised if lexicalised is understood to imply idiosyncratic in meaning to a certain extent; (ii) They are grammatical in all the verbal forms of the paradigm, unlike *benparlat*; (iii) They conform to the Case theoretic considerations on compounding made above, while *benparlat* behaves like an X' in that respect.

In relation to the ill-formed compounds of form *[bo V]_V, I only see a way of accounting for the gap in their generation: these compounds do not exist because the language already has, as part of its listed vocabulary, the words in (26). The words in (26) preempt the generation of (28), this becoming an instance of blocking (cf. Aronoff 1976). New expressions in *ben-* are scarce, but the problem remains as to why they occur instead of compounds in *bo-*.

Finally, with regard to the scarcity of new [A V]_V compounds, except for those with *mal*, I suggest that it seems to have more to do with factors of use than with grammaticality. The newly created (31) do not seem ungrammatical:

- (31) ràpid-cantar
 quick sing-INF 'to sing quickly'

suau-tocar	
soft touch-INF	'to touch softly'

6.1.8. [P V]

Given the assumption made all the way through that prepositions in complement position cannot have their arguments satisfied, this compound structure is predicted to generate ill-formed compounds. This prediction is borne out, as (32) shows.

- | | | |
|------|----------------|---------------|
| (32) | *sota-donar | |
| | under give-INF | '?give under' |
| | *en-agafar | |
| | in take-INF | '?take' |

These "words" cannot even be given an interpretation.

Observe, nevertheless, the existent (33). Other Romance languages present parallel examples: *s'entre-aider* (French) 'to help each other', *entre-lazarse* (Spanish) 'to link with each other'.

- | | | |
|---------|----------------|--|
| (33) a. | entre-cavar | |
| | dig-INF | 'to dig superficially' |
| | *entre-veure | |
| | see-INF | 'to see through the corner of the eye, to glimpse' |
| | entre-badar-se | |
| | open-INF SE | 'to open partly' |
| | entre-girar-se | |
| | turn-INF SE | 'to half turn' |
| | entre-lluir | |
| | shine-INF | 'to be partly apparent' |
| | entre-tallar | |
| | cut-INF | 'to cut superficially' |
| b. | entre-mirar-se | |
| | look-INF REFL | 'to look at each other' |
| | entre-lligar | |
| | tie-INF up | 'to interweave' |
| | entre-matar-se | |
| | kill-INF REFL | 'to kill each other' |

entre-creuar-se
cross-INF REFL 'to intersect, to cross'

contra-dir
say-INF 'to contradict'

Here the first members of the word look like the prepositions *entre* 'between' and *contra* 'against'. I write "look", because they do not always sound like them; compare *contra tu* [kontrətu] 'against you' and *contradir* [kuntrədil] 'to contradict', and *entre mates* [entrəmatəs] 'between bushes' and *s'entrematen* [səntərəmatən] 'they kill each other'. In other cases though, they are pronounced [entrə] and [kontrə] like the prepositions. If *entre* and *contra* are prepositions in (33), in most cases some of their phonological properties have become neutralised. The majority of cases suggest that *entre-* and *contra-* are derivatives; the examples with unreduced vowels are still no conclusive evidence of compoundhood, since there are affixes with strong vowels (i.e. the vowels [e], [o],..., vs. [ə], which appears in unstressed position). Briefly, the phonological evidence allows for the words in (33) to be considered derivatives, with the prefixes *contra-* and *entre-*.

The words in (33) can be divided into two groups with respect to meaning: those in (33a) are interpreted like the base verb with the qualification 'slightly' - or some variant of this pragmatically suited to the verb at hand, thus e.g. *veure* 'to see', *entreveure* 'to see through the corner of the eye'. Those in (33b) are reciprocals of the base verb, e.g. *matar* 'to kill', *entre-matar-se* (with an added REFL clitic) 'to kill each other'⁷. On the other hand, as well as the meaning 'between', *entre* can convey another meaning, 'together', as in (34). (This traces back to vulgar Latin -- cf. Coromines 1982:397). This use of the preposition *entre*, called participative, is found in other Romance languages such as Occitan, Italian, Spanish and French (cf. Rigau 1990).

- (34) Entre tu i jo acabarem la feina.
P you and I we-will-finish the work
'You and I will finish the work together'

The participative meaning of *entre* displayed in (34) is close, though not identical, to the reciprocal meaning in (33b).

In conclusion, both phonological characteristics and semantic properties of the first members of the words in (33) stop us from identifying them, in a synchronic grammar, as prepositions. For these reasons, they will be considered derivatives,

⁷*Matar-se*, with the clitic only, can be interpreted the same way (as reciprocal), but also as reflexive 'to kill themselves'.

instead of compounds⁸.

6.1.9. [NA]

This type of compound is well attested and productive, as shown in (35), where existent (35a), and newly coined compounds (35b) are listed.

- (35) a. *cama-curt*
leg short 'short-legged'
- panxa-content*
belly happy 'laid back'
- cara-llarg*
face long 'unfriendly'
- b. *nas-curt*
nose short 'short-nosed'
- orella-llarg*
ear long 'long-eared'
- coll-tort*
neck crooked 'twisted-necked'

What is of note about all the compounds above is that the noun in first position refers to an inalienable part of an entity. I have already made a proposal for the analysis of inalienability in verbal compounds (see *alatrencar*, section 3.3.2.1.), and I shall explore the possibility of extending that analysis to the data here.

In principle, with the adjective being the head of the compound, the noun in first position can either be the adjective's modifier (conceivably), or its argument. There is no evidence of the noun ever being a semantically unrestricted modifier:

- (36) **girafa-llarg*
giraffe long 'long like a giraffe'
- **gos-rabiüt*
dog rabid 'rabid like a dog'

There are a few cases of an adjective (such as (37)) taking as internal argument the constituent in the first position, as predicted to be allowed if the first position of the compound is an argument position.

⁸Rigau (1990:371) considers these as "cases of morphological incorporation, where *entre* acts as a prefix". In her brief paper, she does not make clear what is meant by "morphological incorporation", though incorporation is currently understood to be syntactic by definition in the works she refers to.

- (37) droga-addicte
drug addict

However, in other occurrences of $[N A]_A$ compounds, the noun is not assigned an internal theta role by the adjective; e.g. in (38) *curt*, unlike *addicte*, has no internal theta role to assign.

- (38) vaig veure _{NP}[aquell _{N'}[personatge _A[camacurt]]]
I-saw that character leg-short
'I saw that short-legged character'

I use coindexation to express inalienability (as done for *alatrençar*). Also, I shall use coindexation to indicate modification (supercoindexation will be used here for clarity). Then (38) is represented as in (39).

- (39) vaig veure [aquell [personatge_i^a [cama_i curt]^a]]

The supercoindexation of a noun with an adjective indicates the kind of predication that holds between the two; this semantic relation manifests itself in agreement of gender and number in Catalan.

The relation of inalienability in (38)-(39) can also be expressed by a maximal projection modifier. Thus (40) is a paraphrase of (38).

- (40) Vaig veure aquell personatge curt de cama.
I-saw that character short of leg
'I saw that short-legged character'

(38) and (40) cannot be interpreted in a way other than one implying inalienability. (41), which does not imply it, is ill-formed.

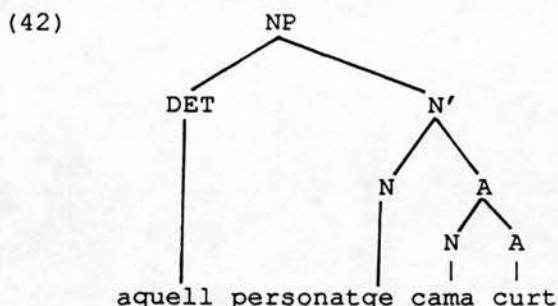
- (41) *Vaig veure aquell personatge curt de pantalons.
I-saw that character short of trousers
'I saw that short-trouser character'

What (38) and (40) have in common is that the nominal denoting the inalienably possessed entity⁹ is c-commanded by the adjective; it also holds that the possessor noun c-commands the possessed noun (i.e. *personatge* c-commands *cama*). It appears, then, that inalienability as construed by Guéron (1984) is applicable. Guéron observes that:

1. The possessor nominal is obligatory.
2. The possessor is in the minimal governing category of the noun denoting the part of an entity.

⁹In what follows I shall designate this as "the possessed nominal", and the nominal denoting the possessor as "the possessor nominal" for short.

3. The inalienably possessed part of an entity is not referential.
4. The two nominals form a chain, and therefore only one theta role may be assigned to them.



As with *alatrencar*, there are examples of compounds where the inalienably possessed is not a part of the body, but more generally a part of something, as expressed by the Non-Distinctness Constraint (cf. 3.3.2.1.1.)¹⁰:

- (43)
- | | | |
|------------|--|------------------------|
| aigua-moll | | |
| water wet | | 'marsh' |
| | | |
| terra-plè | | |
| soil full | | 'terrace (in a field)' |

The problem that arises is: How can the grammar force the inalienable possession reading? So far I have only shown that it allows it.

Comparing again the two prototypical examples, *drogaaddicte* 'drug-addict' and *camacurt* 'short-legged', the inalienable possession reading occurs only in the second case. Consider now that *droga* is the internal argument of *addicte*, and that therefore *addicte* is a two-place predicate (A is addicted to B); on the other hand, *curt* is a one-place predicate (A is short). All the arguments that the adjectival predicate takes are saturated in the phrases *un home drogaaddicte* 'a drug-addict man' and *un home camacurt* 'a short-legged man'. I shall enforce this generalisation by reference to the extended notion of predication. The inalienable possession reading of the one-place cases will be entailed as a result.

In effect, *addicte* in *drogaaddicte* satisfies its internal argument in the first position of the compound; *curt* in *camacurt*, on the other hand, is predicated of the noun *cama*. Both cases are instances of predication as construed in section 3.3.1.2. Alternatively, for the case of *camacurt*, it is possible to adopt Levin and Rappaport's (1986), or Higginbotham's (1985) analyses, according to which adjectives assign an external theta role to the nominal they modify; their approach

¹⁰The examples in (43) are existing compounds, and in contemporary Catalan they are found as nouns only, as result of a category change.

can also be subsumed by my notion of predication. In either case, it is possible to make inalienability follow in adjective compounds if (44) is assumed for Catalan:

- (44) In a compound $[X Y]_Y$, if $X = N$, then X must be an argument of predication.

That is, N in $[N Y]_Y$ in Catalan can only be the argument of predication of a theta role assigner or a modifier. If, in $[N Y]_Y$, $Y = A$ and this adjective is a two-place predicate, the nominal modified in the compound corresponds to the internal argument of the adjective, and the compound adjective as a whole is predicated of a compound-external nominal. (Note that when two arguments need to be saturated, the internal argument is saturated structurally closer to the adjective; hence *un home drogaaddicte* means 'a drug-addict man', not a 'man-addict drug'). If the adjective Y is a one-place predicate, its capacity to predicate being limited to one argument, the nominal modified by the compound adjective and the noun inside the compound will be forced to form a lexical chain: otherwise the compound adjective will fail to predicate in the phrasal structure (and therefore will not be inserted in a phrase), since the predicative capacity of the base adjective will be exhausted. By (44), nouns as modifiers are ungrammatical, because they are not being predicated of in a structure like $[N Y]_Y$: **girafa-llarg*; the pattern is not productive, in fact. Conversely, the lack of a filter such as (44) in English, for example, accounts for the fact that nouns as modifiers are entirely productive, e.g. *fire truck*, *desk stamp*, and so on (cf. the considerations of Lieber (1983) reported in 6.1.1.).

Finally, notice that (44) makes Catalan compounds an instance of semantic Case like *alatrencar* in 6.1.5. Condition (20) signals the noun in a $[N V]_V$ structure as the $THEME_1$ of the verb; here (44) signals it as an argument of predication; being the $THEME_1$ of a verb implies being an argument of predication (of the verb). So satisfaction of the stricter requirement (20) entails satisfaction also of (44).

Before finishing this section I shall briefly look at two more phenomena. First, there are some compound nouns with the structure $[N A]$, which appear to contravene right-headedness and the percolation conventions adopted.

- (45) *un cama-llarg*
 a leg long *'wading bird'*

 un cap-gros
 a big head *'tadpole'*

Note however that the compounds in (45) can also function as adjectives:

- (46) *un noi molt cama-llarg i cap-gros*
 a boy very leg long and head big
 'a long-legged big-headed boy'

(46) indicates that *camallarg* has, together with the specific meaning in (45) 'wading bird', a completely transparent interpretation as adjective, meaning 'long-legged'. This is sufficient evidence to take the examples in (45) to be idiomatic nouns derived by category change from regular adjectives. (I take it that it is not controversial to assume the possibility of category change from adjectives to nouns and so on).

Second, in Chapter Two a dialectal variation was pointed out between continental and Balearic Catalan in connection with agreement. The data are repeated here:

- (47) a. *llengua-llarg* (continental)
 tongue long
 FEM MASC
 'chatterbox'
- b. *llengua-llarga* (Balearic)
 FEM FEM
- c. *un pell-roja* (continental and Balearic)
 a skin red
 MASC FEM FEM
 'an American Indian'

(47c) shows a noun in continental Catalan which follows the pattern that all compound adjectives (and nouns) of this type conform to in Balearic Catalan.

(47a) is what is to be expected in the grammar built up here. Because agreement between two elements and compounding occur complementarily (because of having the same function, namely the fulfilment of the Visibility Condition), no compound-internal agreement is expected between a noun and an adjective. On the basis of this reasoning, (47b) and (47c) are not expected to be grammatical as compounds, because the adjective is marked for gender though this is not an inherent feature of it.

Notice that there are X' in the language with the same word order and gender agreement properties as displayed in (47b,c):

- (48) *té [la [llengua llarga]_{N'}]_{NP}*
 has a longue tongue
 'S/he has a longue tongue'
- Hi ha gent de pell roja.*
 there is people of skin red
 'There's people with red skin'

Therefore, examples (47b) and (47c) can be straightforwardly accommodated in the grammar as N' structures, and they are not going to be considered compounds strictly. In section 6.1.3. examples similar to these were found (e.g. *pocapena* 'little-FEM-sorrow' 'miserable person'), and were treated in the same manner as *llenguallarga*. I shall come back to this kind of constructions in section 6.4.

6.1.10. [V A]

As in sections 6.1.2. and 6.1.6., the argument here is that [V A]_A are ungrammatical because, in non-head position, verbs cannot satisfy their argumental requirements. Thus (49) is accounted for:

- (49) *corre-ràpid
run quick 'quick like running'
- *queixa-displicant
complain unfriendly 'unfriendly as if complaining'
- *treballa-cansat
work tired 'tired of working'

6.1.11. [A A]

[A A] adjectives are quite well represented in the corpora of Catalan compounds, and they are productive. This is to be expected if we consider that the first member of the compound may be a modifier of the second one. Naturally, the first member cannot be an argument of the head adjective, because this is a function that, among X⁰ categories, only nouns can fulfil.

- (50) blau-verd
blue green 'bluish green'
- blau-gris
blue grey 'bluish grey'
- blau-verd-gris
blue green grey 'bluish greenish grey'
- agre-dolç
sour sweet 'sweet and sour'

A semantic restriction seems to hold for the adjective: it must be intersective. An adjective A is an intersective adjective if and only if all sentences of the form 'all A N are N' are valid. *Suposat* 'supposed' and *fals* 'false' are not intersective, and so ??*suposat-blau* 'supposed blue' and ??*fals-daurat* 'false golden' are not good. The meaning of the compounds in (50) is given by the intersection of the meanings of the two elements. We then have a semantic explanation for the ill-formedness of examples such as ??*fals-daurat* and ??*suposat-blau*. Thus whatever the exact nature of "intensional" adjectives such as *fals*, *suposat*, etc. (Dowty et al.

1981:144), it makes no sense to intersect e.g. those things which are 'supposed' which those things that are 'blue'.

The expressions in (51), which might look like those in (50), except for the extra feature of being formed by two (or more) identical adjectives, are nevertheless different, as shown by the inflected forms in (51b).

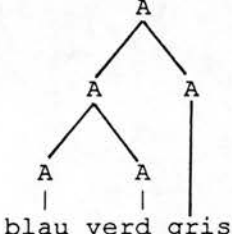
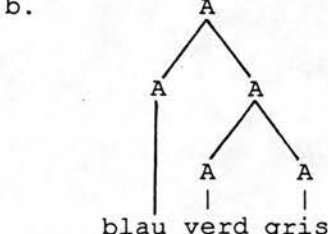
- (51) a. pelut pelut
 hairy hairy
 'very hairy'
- petit petit petit
 small small small
 'very, very small'
- b. uns mitjons petits petits/*petit petits
 a-PL sock-PL small-PL small-PL/small small-PL
 'very small socks'

That is, these sequences of adjectives behave just like any set of adjectives in a phrase (cf. (52)).

- (52) Busco els mitjons blancs finets.
 I-look for the sock-PL white-PL soft-DIM-PL
 'I'm looking for the soft, white little socks'

The right interpretation is also reached straightforwardly from the syntactic structure (one adjective having scope over the other correctly translates into an intensification in (51)).

Cabré and Rigau (1986) raise the issue of the possible structure for a compound of the type of (50) with more than two adjectives. What structure should be chosen for *blau-verd-gris* among those in (53)?

- (53) a. 
 b. 

No matter which one is chosen, the structure has, in this circumstance, no semantic effect: the colour referred to is at the intersection of blue, green and grey. In so far as choosing between (53a) and (53b) has no repercussions for the interpretation of the compound, nor for the determination of its morphosyntactic behaviour, it is unclear how to proceed on this point. Hoeksema (1985) effectively treats such cases as flat n-ary branching (cf. section 6.2.).

6.1.12. [P A]

Apart from the "exocentric" compounds of form [P A] (cf. section 3.3.2.2.), there are no adjectival compounds like this. Again, as in 6.1.4. and 6.1.8. I appeal to the thematic features of Catalan prepositions to rule (54) out: prepositions cannot have their arguments satisfied if they are not in head position.

(54) *davant-fosc
in-front-of dark

*sota-trist
under sad

*damunt-alegre
on merry

6.1.13. [X P]

As discussed above (cf. section 5.1.), prepositions belong to a non-productive (closed) class of words, and no compound prepositions can be newly coined:

(55) a. *teulada-sobre
roof on

*Es teulada-sobre la casa.
it-is roof on the house

Es sobre la teulada.
it-is on the roof

b. *poc-damunt
little on

c. *vola-damunt
fly on

d. *sota-sobre
under on

So, all [X P]_P where X ∈ {N, A, V, P}, are ruled out.

The following existing prepositions, which have (had) this internal structure, are idiosyncratic. As adduced in 1.3., they are part of the grammar because they belong to the lexicon, despite the fact that they do not conform to the generative component of the grammar of Catalan.

- (56)
- | | | |
|----------|--|-----------|
| d-amunt | | |
| of up | | 'on' |
| en-vers | | |
| in wards | | 'towards' |
| mal-grat | | |
| bad will | | 'despite' |
| per a | | |
| for in | | 'for' |
| des de | | |
| from of | | 'from' |

A question that arises in view of the examples in (56) is: How did they ever occur, if the grammar does not leave room for such constructions? Obviously, when they were introduced, the specifications in the grammar need not have been identical to those applying today. Some of the properties of prepositions might have changed in the course of the history of Catalan. In particular, prepositions may have originated in other syntactic categories (as attested in other languages, such as Greek, in which some prepositions originated in nouns). A class that is closed synchronically need not be invariant diachronically.

6.2. Other semantic relations

So far, I have considered two kinds of semantic relation between elements A and B, both predications of A by B.

1. Theta role assignment, when A is an argument of B, i.e. B assigns a theta role to A. This has been discussed in some detail in relation to e.g. the *alatrencar* and the *parallamps* compound types.
2. Modification of A by B, cf. section 3.3.1.

There is at least one other semantic relation exhibited by compounds, exemplified in (57).

- (57)
- | |
|-----------------|
| Austria-Hongria |
| Austria Hungary |
| papà-mamà |
| dad mum |
| 'parents' |

pèrdues-i-guanys
losses and benefits

This is what is called dvandva type in the Sanskrit linguistic tradition, a type which expresses coordination. There are reasons to believe that at least some of the expressions in (57) are compounds, not simply coordinated nouns, given their morphological behaviour:

- (58) a. No volen més Àustria-Hongries a Europa.
no they-want more Austria-Hungaries in Europe
'They don't want more Austro-Hungaries in Europe'
- *No volen més Àustries-Hongries a Europa.
- PL PL

These examples were included among "copulative" compounds by Bloomfield (1933:235), these being headless. He claimed that, of the two elements which make up the compound, there is no way to determine which is the head. However, I see no problem in considering both of them heads (an instance of "splitting heads", cf. Cann 1987). In fact, one of the conditions on the well-formedness of this type of coordination is that the two constituents must share all features -- not only must they belong to the same syntactic category, but they must also have the same thematic properties, as shown in (59); in this way there is no clash of head features when they percolate.

- (59) a. puja-i-baixa pujar [AG, (LOC)]
go-up and go-down baixar [AG, (LOC)]
'action of going up and down'
- estira-i-arrronsa estirar [AG, (TH)]
stretch and shrink arronsa [AG, (TH)]
'bargaining'
- b. *puja-i-veu veure [EXP, (TH)]
go-up and see

However, due to the nominal character of the constructions in (59a), the optional complements are, most commonly, not realised:

- (60) Aquest estira-i-arronsa m'està cansant.
this stretch-and-shrink me is tiring
'I'm tired of this bargaining'
- ?Aquest puja-i-baixa les escales em mareja.
this go up-and-go down the stairs me makes-sick
'This movement up and down the stairs makes me sick'

Here I shall not concentrate any more on the examples on (60) or the examples with the conjunction *i* 'and' in (57) (e.g. *pèrdues-i-guanys*), because, for independent reasons (the presence of CONJ), I shall not consider them compounds *stricto sensu* -- see 6.4.

The coordination compound type cannot be understood as expressing a kind of modification, because *Austria-Hongria*, for example, is not a kind of *Austria* nor of *Hongria*, rather the addition of the two. So (61a) does not imply (61b).

- (61) a. Era a *Austria-Hongria*.
 s/he-was in Austro-Hungary
- b. Era a *Austria*.
 s/he-was in Austria

Other examples, such as (62), though similar to (57) (in that both elements belong to the same syntactic category, etc.), can be interpreted as cases of modification (and so they are in 6.1.).

- (62) blau-verd
 blue green 'bluish green'
- sord-mut
 deaf mute
- agre-dolç
 sour sweet
- nord-est
 North East

Because they are examples of intersective modification, the sentences in (63a) and (64a) imply those in (63b) and (64b), respectively.

- (63) a. Es agre-dolç.
 it-is sour-sweet
 'It is sweet and sour'
- b. Es agre.
 it-is sour
- Es dolç.
 it-is sweet
- (64) a. Es cap al nord-est.
 it-is to the North-East
- b. Es cap al nord.
 it-is to the North
- Es cap a l'est.
 it-is to the East

The distinction that I draw between the types represented by *Austria-Hongria* and *blau-verd* is not generally made in the literature, as far as I am aware. Both types belong to the class of copulative compounds of Bloomfield (1933), called serial compounds by Hoeksema (1985).

Hoeksema exemplifies serial compounds as in (65), of which (65a-b) are called

appositive compounds and characterised as the concatenation of n elements of the same category, with, in principle, no upper bound on n .

- (65) a. root-wit-blau (Dutch)
 red white blue
 b. jazz-rock
 c. de route Amsterdam-Milaan (Dutch)
 the route Amsterdam Milan
 d. Russian-Polish contacts

According to Hoeksema, serial compounds do not display a functor/argument structure, and therefore are not amenable in categorial grammar. Generalisations are hard to make especially for (65c-d), but for the appositive compounds in (65a-b) Hoeksema puts forward the syntactic operations in (66), with their corresponding rules of interpretation, (67).

- (66) $f(N^+) = N$
 where N^+ is any finite nonempty string of nouns

 $f(A^+) = A$
 where A^+ is any finite nonempty string of adjectives
- (67) $||f(N_1, \dots, N_n)|| = ||N_1|| \cap ||N_2|| \cap \dots \cap ||N_n||$
 where $n \geq 1$

 $||f(A_1, \dots, A_n)|| = ||A_1|| \cap ||A_2|| \cap \dots \cap ||A_n||$
 where $n \geq 1$

Contra Hoeksema, I propose that appositive compounds can be handled as functor/argument structures, in particular as cases of modification. These structures may not be motivated syntactically, but they are going to give the desired interpretation (i.e. the interpretation represented in (67) above). Compounds like *Austria-Hongria*, on the other hand, are not interpreted as represented in (67), but rather as represented in (68) (I adopt Hoeksema's notation, here). (68) does not have an adjective counterpart in Catalan: an [A A] sequence is not ambiguous in the way that a [N N] sequence is; there is no union reading for [A A] adjectives. For instance, *blau-verd* cannot be used to designate something with blue and green stripes; for the union reading, *blau i verd* 'blue and green' is the only adequate expression.

- (68) $||f(N_1, \dots, N_n)|| = ||N_1|| \cup ||N_2|| \cup \dots \cup ||N_n||$
 where $n \geq 1$

The compounds whose interpretation is represented by (68) truly cannot be subsumed by a functor/argument structure, and will be assigned a coordination structure. By this, I give some indirect support for the otherwise unmotivated

binary structure assigned to *blau-verd* and the like: I make use of two syntactic structures by associating them with two different interpretations¹¹. This means that I can maintain a one-to-one correspondence between the assigned syntactic structure and the semantics for compounds such as *Austria-Hongria* and *blau-verd*. (Notice, as well, that if the *blau-verd* type were an instance of coordination, we would be left with an unaccounted gap in the lexicon, that of the [A A]_A headed compounds; this problem does not arise here).

6.3. A remaining issue: The case of *enriure*'s

Fabra (1956) draws attention to the complex words illustrated by the following examples, which he considers to be compounds.

- (69) a. Va enriure's del seu pare.
s/he en-laughed-SE of-DET his/her father
'S/he laughed at her/his father'
- b. Va emportar-se una capsa.
s/he en-took-SE a box
'S/he took a box'

The question arising is whether *en* is a derivational affix, or a clitic (as Fabra claims). The clitic *en* of Catalan (*en* in French, *ne* in Italian) appears in complementary distribution with a NP category in the manner illustrated in (70).

- (70) a. Parlo de la Montserrat.
I-talk of DET Montserrat
'I'm talking about Montserrat'
- En parlo.
EN I-talk
'I'm talking about her'
- *En parlo de la Montserrat.
EN I-talk of DET Montserrat
- b. Compraré una dotzena d'ous.
I-will-buy a dozen of eggs
'I'll buy a dozen eggs'
- En compraré una dotzena.
NE I-will-buy a dozen
'I'll buy a dozen of them'
- *En compraré una dotzena d'ous.
NE I-will-buy a dozen of eggs

Clitics are generally regarded as referential maximal projections; see e.g. Radford

¹¹Indeed one of the associations, namely the intersective one, is independently motivated.

1981 for a short comment on the matter, and Borer 1986 for discussion. Kayne (1989:245), by contrast, considers clitics to be of category X^0 . According to the latter view, it would be possible for a clitic to be a part of a compound. If the clitic *en* is a phrasal projection, it is excluded as a member of a compound.

If *en* in the following is a clitic, then the examples show double occurrence of that clitic form:

- (71) a. Va enriure-se'n.
 s/he en-laughed-SE EN
 'S/he laughed at him/her/it'
- b. Va emportar-se'n una.
 s/he en-took-SE EN one
 'S/he took one'

This gives a strong argument to exclude the analysis of *enriure*'s and *emportar-se* in terms of a clitic, since no other repeated occurrence of the same clitic on the same verb is attested (to my knowledge this is not only so for Catalan, but also for all languages with clitics).

- (72) a. L'hi dono a casa.
 it DAT I-give at home
 'I give it to him/her at home'
- b. A la fotografia hi dono un llibre a la Marta.
 in the photograph there I-give a book to DET Marta
 'In the photograph I'm giving a book to Marta'
- c. *L'hi hi dono.
 it DAT there I-give
 'I give it to him/her there'

(72a-b) show the independent cliticisation of elements fulfilling GOAL and LOCATIVE theta roles, both of them (on this occasion) cliticising in the clitic *hi*. Example (72a) also shows the cooccurrence of two different clitics. Example (72c), however, shows, that GOAL and LOCATIVE arguments cannot both cliticise in *hi* at the same time.

We have to account for the ungrammaticality of (72c) by establishing that a form can appear only once as clitic of a verb, regardless of the theta role that it is assigned. The cooccurrence of two morphologically identical clitics even with different theta roles must be blocked. This constraint can be derived from Case theory: if clitics are assigned Case by the verb, and each verb can assign Case to one argument of each type (ACC, etc.) at most, no two clitics can be assigned the same Case by the same verb. An analysis of *enriure*'s containing a clitic is in contradiction with these observations.

Analysing *en* as a prefix has, on the other hand, the following consequences. *En-*

as a prefix attaches to a verb in its inherently pronominal form, as in the examples in (73). The number of verbs of the kind seems to be rather small.

- (73) *em-portar-se*
 take-INF-SE 'to take with'
- en-dur-se*
 take-INF-SE (same meaning)
- en-recordar-se*
 remember-INF-SE 'to recall'

The non-prefixed **portar-se*, **dur-se* are ungrammatical (the first only so with the intended meaning). The verbs in (73) have at their bases *portar*, *dur*, *recordar*, not all the verbs like them have a counterpart of the form *en-V-se*, and I see no way of predicting when they do indeed have it.

- (74) a. *transportar d'aquí a allà*
 'to transport from here to there'
- *en-transportar-se*
- b. *oblidar alguna cosa*
 'forget something'
- *en-oblidar-se*

Therefore, the verbs in (73) do not constitute a class for any purposes other than having *en-V-se* counterparts; these pairs must, then, be listed in the lexicon.

Still, the first two examples in (73) are enlightening with regard to the origin of such verbs. Compare the first example in (73) with (75).

- (75) a. *Porta el cotxe al garatge.*
 s/he-takes the car to-the garage
- b. *S'emporta el cotxe.*
 SE en-takes the car
 'S/he takes the car with him/her'

Emportar-se has added to the meaning of *portar* the specification 'with her/him, away from here', which is the meaning that a [*de* NP] complement would add (the [*de* NP] complements are the ones which cliticise in *en*). This pattern is far from general¹², but gives us an idea as to what might have been the origin of the word, which involves the clitic *en*. However, at no point in history is there evidence that the two *en* forms coexisted as clitics. As shown above, there are reasons to believe that they have never coexisted as clitics, though the affix *en-* may well be a fossilised one, syntactically inert.

¹²Though the same examples are found in French: *porter*, *apporter*, *emporter*,..

Recall finally the examples of Fabra (1956) from Chapter Two, repeated here for convenience:

- (76) a. *Va riure's del seu pare moderadament.*
 s/he laughed SE of-DET his/her father moderately
 'S/he laughed at her/his father moderately'
- b. **Va enriure's del seu pare moderadament.*
 s/he-en-laughed-SE of-the her/his father moderately
- c. *Capses? Va emportar-se'n quatre.*
 Boxes? s/he-en-took-SE EN four
 'Boxes? S/he took four of them'

Of (76b), note that I find the example grammatical, as is expected. If Fabra found it ungrammatical it is presumably because he understood it as a cliticisation of (76a), rather than as a derivative with an *en-* prefix. There was no verb *enriure's* for Fabra, then. Of (76c), it seems that Fabra did understand the complex word *emportar-se* as a derivative.

6.4. Lexicalised expressions

All along I have argued for a rather strict notion of compounding. Many expressions that in the literature are considered compounds are not analysed as such here. If my claim is that they are not compounds, I must provide an alternative analysis for them. This is the purpose of this section.

The criteria on which I have decided to exclude certain expressions are largely morphosyntactic and syntactic, related to the subtheory having to do with the language-specific expression of meaning, i.e. Case theory. It was not on the grounds of theta theory, which has to do with more universal semantic issues. That is, thematic relations, I argue, can hold equally within or outside compounds¹³, but they are grammaticalised differently through compounding or through phrase formation.

The expressions in (77) are not considered compounds proper (some of them appear in Chapters Two to Five and are repeated here for convenience).

- (77) *pa de pessic*
 bread of pinch 'sponge cake'

¹³This does not exclude some semantic difference between arguments as phrases or within compounds (e.g. the difference generic/non-generic).

aigua-ardent water burning	'liquor'
mitja-nit half night	'mid-night'
gos-llop dog wolf	'wolf-hound'
pruna clàudia plum Claudia	'greengage'
cap-i-cua head and tail	'palindrome number'
a-les-hores in the hours	'then'

Because my rendering of compounding cannot generate them, the following strong claim is made: that there is, in a grammar based on principles and parameters such as GB, a way of generating them which of course does not involve compounding. Without having to postulate some new, unprecedented type of word formation, the above requirement is satisfied. The expressions in (77) are formed by the standard means of phrase structure generation, with a minor addition, namely lexicalisation. To prove the point, consider the phrases in (78), as freely generated in the syntax.

- (78)
- | | |
|---|------------------|
| pa d'ordi
bread of barley | 'barley bread' |
| aigua bullent
water boiling | 'boiling water' |
| mitja taronja
half-FEM orange | 'half an orange' |
| tinc [gat i gos]
I-have cat and dog
'I have both cat and dog' | |
| ho faré [a les vacances]
it I-will-do in the-PL holiday-PL
'I'll do it during the holidays' | |

Like X' and phrases in Catalan, the expressions in (77) have the following characteristics:

1. They are mostly head initial;
2. The nominals occurring in them need to be licensed by the Case Filter;
3. They include non-inherent inflectional markers in positions other than the periphery of the X'/XP.

These three conditions are not satisfied by Catalan compounds. Let us review them one at a time.

I have defended the position that the grammar of Catalan must include a percolation convention to the effect that compounds are head final¹⁴. It seems harder to generalise on the position of heads in phrases, but most phrases are head initial in Catalan. (Subjects being quite often superficially sentence initial, they are an exception to the above statement unless they are posited to be post-verbal at D-structure. There are in fact some proposals along these lines, cf. Bonet 1989, reported in Rígau 1990). The only exception to this in (78) is *mitja taronja*, where the noun is preceded by *mitja* 'half'. If *mitja* is a determiner, then it is possible to adopt the theory according to which determiners are heads of DP (cf. Szabolcsi 1987 and the references there), and so this word order is the expected one:



If *mitja* is an adjective (as suggested by *La mitja taronja que m'he menjat era boníssima* 'the half orange that I have eaten was very good', where *mitja* and a DET cooccur), we have to group this adjective with a few others that are pre-nominal (e.g. *pobre* in *pobre fill* 'poor child', *mal* in *mal home* 'bad man'). In either case, the expressions in (77) and (78) have the same word order, which is the point which needs to be made here. This account carries through to *pocapena*, which has been encountered in 6.1.3.

In connection with the Case Filter, the presence of a Case marker such as *de* 'of' in *pa de pessic* indicates that the Case Filter applies here, unlike inside compounds (thus **pa pessic*, the same as **pa ordi*). In contrast to this, a few examples have to be dealt with, e.g. *gos llop* 'wolf dog', and *carrer Muntaner* 'Muntaner street'. These are sequences of a noun followed by another noun, where the second, the non-head, is not assigned Case by *de* (*carrer de Muntaner* is acceptable, but **gos de llop* is certainly not). The second noun is, rather than a complement, a modifier of the first one. This kind of modification is quite limited: only a few cases are found, some of which date back to Latin, where the second noun was truly Case-marked by a morphological case marker. Other examples are given in (80):

¹⁴No evidence has been found of a construction which could only be analysed as compound and was left-headed, the *parallamps* type aside.

(80)	paper moneda paper coin	'bill'
	vagó restaurant wagon restaurant	'restaurant car'
	pasta dents paste tooth-PL	'tooth-paste'
	tintura iode tincture iodine	'iodine'

Because the restrictions derived from the Case Filter do not hold in compounding, this type of modification of a noun by a noun would be predicted to be potentially more productive in compounding, while irregular among phrases, were it not for the considerations in section 6.1.9. on nouns being the objects of predication in compounds. In fact, $[N N]_N$ are rare and unproductive. The only productive $[N N]_N$ are those naming streets, etc., where a strictly analogical operation (as opposed to syntactic generation) might be at work¹⁵.

Prepositions generated for the satisfaction of the Case Filter or with full semantic content are found in (81), as well as some examples in (77).

(81)	tauleta de nit table-DIM of night	'bed-side table'
	esperit de vi spirit of wine	'alcohol'
	cuc de seda worm of silk	'silk-worm'
	impost sobre la renda tax on the income	'income tax'
	cursa contra rellotge race against watch	'timed race'

Though considered compounds by many (cf. Ferrater 1981b), I treat them as lexicalised X' as well, mainly for the Case theoretic reasons adduced.

Finally, in connection with inflection, most of the expressions here inflect the same as the rest of X'. Compare (82a) and (82b).

(82)	a. pa d'ordi	pans d'ordi
		PL
	'barley bread/s'	

¹⁵For the specific problems of proper names in apposition (e.g. *carrer Muntaner*), see Burton-Roberts 1975, and Matthews 1981.

pruna vermella	prunes vermelles
'red plum/s'	PL PL
màquina per rentar	màquines per rentar
'machine to be washed'	PL
b. pa de pessic	pans de pessic
'spongecake/s'	PL
pruna clàudia	prunes clàudies
'greengage/s'	PL PL
màquina de rentar	màquines de rentar
'washing machine/s'	PL

However, some forms have a behaviour diverging from the predicted one, which is displayed in (82). These are of two kinds, (83) and (84b) (to be compared with (84a)).

(83) gos-llop

gossos-llop
PL

?gossos-llops
PL PL

peix-martell

peixos-martell
PL
'hammer fish/es'

?peixos-martells
PL PL

(84) a. aigua bullent

aigües bullents
PL PL

*aigua bullents
PL

llengua llarga

llengües llargues
PL PL

*llengua llargues
PL

gat i gos

gats i gossos
PL PL

mal de cap

mals de cap
PL
'headache'

b. <i>aigua-ardent</i>	
* <i>aigües-ardents</i>	<i>aigua-ardents</i>
PL PL	PL
<i>cap-i-cua</i>	
* <i>caps-i-cues</i>	<i>cap-i-cues</i>
PL PL	PL
<i>mal-de-cap</i>	
* <i>mals-de-cap</i>	<i>mal-de-caps</i>
PL	PL
'preoccupations'	

So, the normal agreement mechanisms are not able to handle some of these complex expressions if they are of structure $[N A]_N$ (or $[N N]_N$). In (83) the non-head noun is invariable (like a handful of adjectives are). In (84b) the whole expression is inflected like a simple noun. Here is where lexicalisation seems to play a role.

Notice that these examples have been categorised as one-bar projection nominals (leaving aside some of them, which are phrases: $[a [les [hores]]_{NP}]_{PP}$). This is because they cannot be maximal projections, since they do not include a DET nor exclude the presence of one, as shown in (84).

- (85) *embolica* $_{NP}$ $[el [pa de pessic]]$
 wrap-up the bread of pinch
 'Wrap up the sponge cake!'
- vaig sentir udolar* $_{NP}$ $[un [gos-llop]]$
 I-heard howl-INF a dog-wolf
 'I heard a wolf-hound howling'

Second, they cannot be straightforward X^0 projections because then they would be compounds in the narrow sense of the word, and they are not for the reasons adduced above (that is, because they pattern differently from compounds in many respects).

Having rejected their categorisation as NP (*aleshores* and the like aside) and N, it follows that they must be N'. Yet those that inflect like words (e.g. *aiguardents*) indicate thereby that they are in fact N. This is of course the last stage in lexicalisation, and it is not regular or predictable (thus the expressions like *aiguardent* must be listed as nouns and there is no reason to believe that their internal structure will remain transparent to speakers -- e.g. *salfumant* 'hydrochloric acid', which was of the same type, is now a simple phonological word and its structure hardly recognised).

The forms in (80), (81) and (82b), which behave like X' for inflectional matters are also lexicalised. The crucial difference between these N's and those in (78) and

(82a) is that they are peculiar in meaning; so for instance a *màquina de rentar* designates a washing machine, a machine to wash clothes, not cars nor dishes. There is nothing in the generative component of the grammar which could prevent us from using the same expression *màquina de rentar* to refer to a dishwasher, but the fact is that it does not, and the speaker uses it to refer to washing machines only. Briefly, these expressions are associated with a kind of entity in particular, beyond the determination of the grammar on its own. As a consequence, the members of these N's have lost their freedom to refer independently, and have become generic. So, for instance, a *tauleta de nit*, literally 'little-table of night', 'bed-side table', is not a table used during one or several nights in particular, but rather associated with a generic night.

This process of lexicalisation has the following syntactic effects: (i) The constituents of the expression cannot be modified (see (86a)); (ii) They do not allow interruption by other constituents (see (86b)).

- (86) a. *tauleta [de [nit d'estiu]]
 table-DIM of night of summer
- *tauleta de nit que no puc dormir
 table-DIM of night that not I-can sleep
 'night-table that I cannot sleep'
- b. pa de pessic del dia
 bread of pinch of-the day
 'fresh spongecake'
- *pa del dia de pessic

This is the only difference between these N's and the ones in (78) and (82a), and can be entirely attributed to them being lexicalised. The same behaviour can be noted in lexicalised complex predicates (see section 4.3.), which are another instance of lexicalised X', verbal instead of nominal. (I consider a complex predicate to be a V', as in Jayaseelan 1988, if not a VP, given that, in general, I take the constituent formed by a verb and its complement not to be an X⁰ projection). The effect of lexicalisation in complex predicates stops cliticisation of the complement, as shown in (87).

- (87) a. Sempre estàs somiant truites.
 always you-are dreaming trouts
 'You're always dreaming'
- b. Sempre estàs somiant històries de por.
 always you-are dreaming stories of fear
 'You're always dreaming horror stories'
- c. Sempre n'estàs somiant.
 always EN you-are dreaming
 * as cliticisation of (a)

If lexicalised X's have been traced for $X \in \{V, N\}$, one expects to find them for $X = A$ as well, other factors being equal. This expectation is borne out, as shown by (88)¹⁶.

(88) *curt de gambals*
short of mental-space 'blockhead'

dur de pelar
hard of peel-INF 'stubborn'

It is not surprising to discover that lexicalisation is attested for all productive major categories, and not only for an arbitrary subset of them.

What is exceptional about these formations is that they are idiosyncratic in meaning, but that they are entirely regular in the structural mechanism underlying them, namely concatenation under the principles that govern all syntactic structures (the principles of Case theory, X-bar theory, etc.). This implies that expressions like those in (81) are freely generated, and it is a matter of use and time for them to become lexicalised -- lexicalisation remaining an unpredictable factor. This theoretical approach leaves room for new lexicalised X's, which are indeed coined by speakers in their every day use of the language.

Finally, notice that the generic quality of the constituents of the elements inside the lexicalised N' is also found in the constituents of compounds. (The fact that there is, in Catalan, compound-internal theta role assignment does not disprove this point, precisely because the noun in $[N V]_V$ compounds can remain generic in my analysis though it is coindexed with another nominal which is referential).

Given the structure assigned to the lexicalised X', a principled classification of simple and complex forms (including compounds proper) emerges. This is summarised in (89).

¹⁶Some lexicalised PP are also found, e.g. (56), in section 6.1.13., though these seem to belong to the class of Ps and not PP any longer.

(89)

XP	lexicalised
X'	lexicalised X' $X \in \{N, A, V\}$
	called complex predicates when $X = V$
X	simple {a, b, c, ...}
	complex
	compounds $[X, Y]_X$
	derivatives $[aff, Y]_X$
	(+learned complex words)

Of these, some are necessarily idiomatic: the lexicalised X', including lexicalised complex predicates. Other complex expressions may or may not be straightforwardly compositional, as has already been shown.

A last pending issue is: Why are the expressions in 6.2. like *Austria-Hongria* among compounds instead of among lexicalised X's? The fact that they pluralise like simple words (e.g. *Austria-Hongries*) could be accommodated by having this example undergoing the last stage in lexicalisation. I shall adduce three reasons which suggest that *Austria-Hongria* and the like are better treated as compounds. First, there is no evidence of lexicalised $[N N]_N$ interpreted as coordinations at any stage of lexicalisation. This means that no independent evidence can be brought to support *Austria-Hongria* to be an N'. Second, there is no reason to prefer an irregular mechanism (namely lexicalisation) to generate what a free application of the existing principles can already generate. Most importantly, there are lexicalised N' which are not interpreted as union; analysing *Austria-Hongria* and the like together with them would make the semantic distinction between the two groups impossible.

6.5. Catalan compounding in the context of Romance

A topic that I still have not addressed is the parallelism between Catalan and other Romance languages as far as compounding is concerned. This chapter starts with a quotation from Di Sciullo and Williams where they assert that there are no compounds in French and Spanish. I cannot do full justice here to complex expressions in these Romance languages, and shall simply make a few points. While I do not follow Di Sciullo and Williams in their classification of complex words (which separates "syntactic words" from "syntactic phrases" -- see above), it is true that one of the words which appears to be unequivocally a Catalan compound, *alatrencar*, does not have an equivalent in French or Spanish. Native

speakers inform me that this is also the case in Italian. Compare (15) in 6.1.5. with (90).

- (90) a. *corazón-robar (Spanish)
 heart steal-INF

 *corazón-secar
 heart dry-INF

 *cuello-romper
 neck break-INF

 *nariz-partir
 nose split-INF
- b. *coeur-prendre (French)
 heart take-INF

 *coeur-secher
 heart dry-INF

 *cou-rompre
 neck break-INF

 *nez-fendre
 nose split-INF

To merely describe the facts (on the basis of the limited data at hand) we have to say that neither in French nor in Spanish are theta roles realised inside right-headed compounds. Since they are in Catalan, we are confronted with a variation across languages ultimately reducible to a difference in parameterisation.¹⁷ Suppose the parameter in question is formulated as in (91).

- (91) Theta roles {can/cannot} be satisfied inside a
 right-headed compound.

The value for this parameter is the affirmative one for Catalan (and English), the negative one for French and Spanish. We know that the positive setting of the parameter generates expressions that the negative setting does not allow (such as (90)). Assuming the Subset Principle (cf. Berwick 1985), it follows that the negative setting of (91) is the unmarked one¹⁷.

In contrast with Spanish and French, forms like those in (90) were well-formed in Latin (cf. Duarte and Alsina 1986:156). Darmesteter (1875:139ff) gives quite a number of Latin forms like these, as well as and their French fossilised decedents, and claims that the Latin construction was not a compound, but a phrase.

¹⁷The remaining issue is now: How is the parameter set for its marked option? Positive evidence is presumably a way of setting it positively, but compound-internal theta role assignment does not seem to belong to the primary data that a child is exposed to. I shall leave this inquiry open.

- (92) manu-mittēre
 hand give-away-INF 'to give freedom (to a slave)'
- venum-dare
 sale give-INF 'to sell'

If, despite Darmesteter, the expressions in (92) were truly compounds, somewhere along the line of the history between Latin and contemporary French and Spanish this kind of compound got lost, while it did not in Catalan. Catalan speakers, however, do not make as creative a use of these compounds as they probably did in the past (besides, this kind of compound is not learned, but colloquial, and so new formations do not arise in specialised vocabularies, academic contexts, and so on).

On the other hand, new objects are often designated by X'/XP, which unlike most compounds are left-headed, and become lexicalised. So, overall, the trend seems to be, in the present use of the language, to resort to left-headed syntactic structures (including X' and compounds of the *parallamps* type, section 3.3.2.2.) rather than right-headed compounds. This may ultimately affect the grammar of Catalan (and may already have had its impact in those of French and Spanish), but for the time being is a fact of use, so that speakers (still) have intuitions about verbal and non-verbal compounds. It is on these that I have based my work.

Chapter Seven

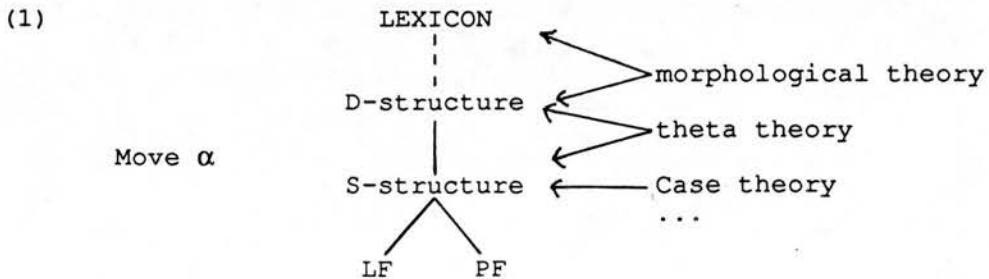
Conclusion

In this chapter I shall outline a morphological theory circumscribing the above treatment of Catalan compounds, as well as indicate some topics for further inquiry which this thesis suggests.

7.1. Morphological theory

The morphological theory here is not unique to this thesis, and though it has not been thoroughly worked out elsewhere, sketches of it can be found in e.g. Shibatani and Kageyama 1988, Kageyama 1989, and Baker 1988a, 1988c. The term used in these works is "morphology theory", instead of which I use "morphological theory", the latter being analogous with terms such as "syntactic theory". By embracing their conception of the theory of morphology, I do not, of course, commit myself to other aspects of the analyses presented in these references.

In a modular theory of grammar such as GB, morphological theory can be envisaged as a subtheory, to be included among binding theory, Case theory, theta theory, and so on. This can be represented as in (1):



The defining characteristic of morphological theory is that it includes all the principles of word well-formedness. Baker (1988a:428-9) writes:

"The rules and principles of morphology are not a subpart of any particular level of the grammar, such as the lexicon or the level of Phonological Form. Instead, they constitute their own semi-independent component of the grammar, and as such, they may constrain representations at any or all levels of description. (...) The domain of morphology theory is the structure of X^0 categories, just as the domain of X-bar theory is the structure of X' and XP level categories."

As underlined by Baker (1988a:430), this view of the relation between syntax and morphology is inconsistent with the model put forward by Lexical Phonology (cf. Kiparsky 1985 and Halle and Mohanan 1985), where syntax and morphology are thought of as sets of independent levels of representation.

Some principles belonging to this morphological theory have been proposed in previous chapters; these are reviewed below, together with constraints that have been postulated elsewhere. Before that, I shall recapitulate on the motivation for a syntactic approach to word formation, as illustrated by the compounds of Catalan.

The argument of this thesis has been that there are three kinds of Catalan compounds:

1. right-headed compounds
2. left-headed compounds of the *parallamps* type, which involve an empty category, pro_N
3. coordinated compounds, where coordination is interpreted as union.

Of these, the second group could be considered peripheral in the grammar of Catalan (and other languages which also allow it, such as Spanish, French, and Italian), as was considered by Di Sciullo and Williams 1987.

What is missing for a proper treatment of the third group is an explanation of why it is restricted to noun compounds. Still, the compounds of this kind can be accounted for by a grammar based on principles and parameters, in which principles and rules (i.e. Move α) apply freely. Briefly, we can account for the grammaticality or ungrammaticality of all the compounds of Catalan by application of the theory of grammar as designed to account for sentential well-formedness.

7.1.1. The Projection Principle

Of all the principles satisfied by Catalan compounds, the one which is the core of GB syntax, namely the Projection Principle, is most central here. Effectively, the analyses proposed rest on the premise that the properties of lexical items as they are encoded in their lexical entries remain unaltered through the derivation; for each case examined, I have been able to maintain the homogeneity of the lexical properties of the lexical items at the base of compounds. For example, for the analysis of *alatrencar*, it has been unnecessary to postulate a lexical entry for *trencar* in *alatrencar* different from the one in free occurrences of *trencar*. The lexical requirements of *trencar* are always the same, and are appropriately satisfied whether it occurs as a constituent of a compound or not. Among other properties of lexical items, I have proposed that its transitivity valence (represented here by the [+/- Case] feature) must be satisfied in a syntactic representation. So, if a verb or a preposition is a Case assigner, its Case-assigning capacity must be realised (though not necessarily in an overt complement).

As reported in section 3.1.2.2., Borer (1984) considers that the crucial distinction between word formation in the lexicon and word formation in the syntax is that it is required of the second that it respect the PrPr, while this condition does not hold for the first. This idea leads her to conclude that traditional inflectional phenomena should take place in the syntax. My conclusion is that this is also the case for Catalan compounds. The same conclusion is reached by e.g. Roeper (1988) for English deverbal compounds¹.

Borer perceives her theory as a new formulation of the Lexicalist Hypothesis, stripped of spurious generalisations. According to her, the Lexicalist Hypothesis establishes that word formation is autonomous from the syntax, except for those inflectional phenomena which respect the PrPr. Permitting syntactic principles to have access to the constituents of all complex words (as done here for compounds) constitutes, though, a plain rejection of the Lexicalist Hypothesis in any of its non-trivial formulations. In a constraint-based grammar, this move allows for a more general application of syntactic principles. Also, it leaves room for a new set of well-formedness conditions (those which form morphological theory), possibly applying -- as all other conditions -- at more than one level of representation. Hence, the rejection of the Lexicalist Hypothesis makes possible the construction of a theory of morphology in tune with a modular grammar.

¹In the context of Dowty's (1979) classification (see section 3.3.1.), the analysis postulated for Catalan compounding makes it a morphological operation by syntactic rule. In contrast to this, an application of Dowty's vocabulary to compounding can be found in Sugioka 1984, where Japanese compounding is considered to be a morphological operation by lexical rule.

As already pointed out, word formation in the syntax is not necessarily word formation via transformation -- despite the fact that these two notions are often conflated in the literature, probably due to the fact that most analyses of compounding in the syntax have involved the application of transformations (from Lees 1960 to Roeper and Siegel 1978 and Baker 1988a). As for all syntactic structures for whose analysis no transformation is postulated, the only basic rule applying here is concatenation of lexical items as they are projected from the lexicon.

Unlike other treatments of compounding such as those of Ferrater (1981b) and Mascaró (1986) for Catalan, the one here does not take sentence and phrase formation as more basic than compound formation. That is, compounding is conceived as an independent means to express meaning, not subsidiary to other means of grammaticalisation. Principles of well-formedness apply to compounds in the same way that they apply to other structures, i.e. when their structural conditions are met; compounds do not enjoy a special status in the grammar.

Specifically, theta theory has been applied in Chapter Three, and the Theta Criterion shown not to be violated. The application of Case theory to the domain of compounds has been dealt with in Chapter Four. The principles of government theory have also been demonstrated to hold compound internally, in Chapter Five. Finally, the binding principles A, B and C are also respected and correctly warrant the interaction of compounding and inalienability found in Catalan, as discussed in section 3.3.2.

7.1.2. Some principles of morphological theory

Within morphological theory, one of the principles which has been assumed in this thesis is the following, as formulated in Baker 1988a:140:

- (2) Stray Affix Filter
*X if X is a lexical item whose morphological
subcategorisation frame is not satisfied at S-structure.

This principle enforces that affixes must attach to words and it is what forces the V-to-INFL movement analysis as summarised in section 5.1.1.3. This principle applies vacuously to compounds, in so far as the basic constituents of these are not affixes.

More relevant for compounding is the following principle, which blocks movement out of a word (a filter which follows from the Lexicalist Hypothesis, when this is maintained).

(3) $*[X^0 \dots t_i \dots]$

In fact, no movement out of a compound, with the consequent remaining trace inside it, has been argued for.

Another principle proposed by Baker (1988a:72) rules out constituents of X' or phrasal projection in words:

(4) $*X^0$
 $\quad |$
 $\quad X^n$
 where n is greater than zero

Yet, (4) is probably too strong for languages such as Dutch and German where compounds appear to have constituents which are not minimal projections (Baker himself points this out, though he finds (4) appropriate for English). French examples such as *trompe-l'oeil* (section 3.3.2.2.) are also problematic for the constraint encoded in (4). Hoeksema (1988) bears on this issue, and observes that maximal projections, if they occur in compounds, are limited to non-head position. Though I shall not discuss this issue in any depth of detail, accommodating Hoeksema's remark, the following parameter of morphological theory can be proposed:

(5) In a structure $[Y^n, X^0]_{X^0}$,
 n {may/may not} be bigger than zero.
 may not: Catalan, Spanish, ...
 may: Dutch, ...

This parameter expresses that, in compounds, the complement of a head cannot be a phrase or an X' projection, though in other circumstances, complements are taken to be maximal projections. This constitutes a conflict between X-bar syntax and compounding. On the other hand, that a structure with an X^0 head is itself an X^0 projection is not a violation of the X-bar conventions; the X-bar theoretic conditions allow the projection α of a category β to be of the same bar-level as β or greater by one level. When the bar-level of the daughter and the mother are the same, we have what has been called "base-generated adjunction structures".

Language-specific filters of morphological theory have been motivated in sections 6.1.5. and 6.1.9., and are as follows:

- (6) In the configuration $[N V]_V$, N can only satisfy a TH_1 (in Catalan).
 (7) In a compound $[X Y]_V$, if $X = N$, then X must be an argument of predication (in Catalan).

(6) is a particular case of (7).

7.1.3. The characterisation of compounds

From these considerations of X-bar theoretic notions, as well as those on Case theory in section 4.2., a definitional characterisation of compounds emerges:

- Compounds are $[X^0, Y^n]_{X^0}$ structures, where $n \in \{0,1,2\}$ as set parametrically, and where X, Y are limited to belonging to major categories, and are not affixal.
- Compounding constitutes in itself an expression of g-Case, so that nominal complements in compounds are identified (cf. the Visibility Condition) in virtue of their concatenation to an X^0 category to form another X^0 category.

Notice that there is no need to stipulate that compound prepositions are ungrammatical, assuming an independent statement of the closed character of this syntactic category required for other purposes as well. Hence, even if prepositions belong to a major category, the exclusion of compound prepositions need not be part of a characterisation of compounds.

A semantic feature of compounds is the generic, non-referential quality of the nouns which take part in them. This property has for long been associated with compounds in the linguistic tradition. Recently, in the context of X-bar theory, a relation between referentiality and maximal projection has been advocated (cf. Sproat 1985, Roeper 1988, Williams 1989). Though the correlation holds for the data explored here, the non-referential quality of compound-internal nouns has not been taken as a premise in my analysis; rather, it *follows* from it (see sections 6.1.5. and 6.1.9.). Nevertheless, it is worth underlining that the NPs which are part of lexicalised expressions do not retain their referring capacity, as exemplified by *nit* 'night' in *tauleta de nit* 'bed-side table'. This suggests that more than just X-bar theoretic considerations is at play in connection with referentiality. It is tempting to think that this semantic feature of nouns in compounds is derivable from other features (e.g. inalienability, as well as X-bar conditions), and therefore I do not include it among the definitional -- i.e. sufficient and necessary -- conditions of compoundhood.

The two definitional characteristics above are certainly theory internal. This is desirable, in so far as whether an expression is a compound or not is not given, but a theoretical issue. Native speakers do not know about compounds, and when questioning them recourse to examples is indispensable. This is not surprising:

"compound" belongs to the linguist's vocabulary, in the same way that NP does. Yet my claim is that the expressions I designate with the term "compound" constitute a natural class, that is, a class that allows us to make predictions on the linguistic behaviour of its members.

7.2. Topics for future research

With respect to the Principles and Parameters grammar invoked here, I suggest four topics which deserve further attention. First, several questions have been left unanswered when dealing with the *parallamps* compound type, particularly on the licensing of the empty category pro_N . Similar questions arise with Rizzi's (1986) *pro* empty object in Italian (see also empty objects in Brazilian Portuguese, cf. Negrão 1990). Thus *pro* categories in Romance are relatively poorly understood. As well as these, empty categories in morphology (i.e. as constituents of words) are generally treated in a way which differs considerably from what is usual in studies on phrase structure, and in a manner which appears very ad hoc. A comprehensive appraisal of empty categories (i.e. in- and out-side X^0 categories) in a broad perspective is necessary.

Second, it has been assumed (as in current work in GB) that affixes are characterised, in the lexicon, for the same features as free forms; therefore, they can belong to a syntactic category, can have theta-grids, etc. At the same time, they have subcategorisation frames that indicate that (i) They are affixes, i.e. they require a sister category, and (ii) Their sister must belong to a particular syntactic category (or one of several syntactic categories). While (i) is necessarily included in the lexical entry of an item, since it is unpredictable, it might not be so for (ii), the specification of the syntactic category of a sister. If this information can be dispensed with for non-affixes (in favour of thematic and Case theoretic information, together with Canonical Structural Realisation statements), it would be interesting to see whether the same strategy can be used for affixes, and, if not, for what reason.

Third, a study along the lines of a GB grammar could be designed for affixation. Such an enterprise is part of the project which this thesis belongs to, namely a Principles and Parameters account of word building. (In fact, the research on "inheritance" of thematic properties in derivatives responds to a similar program -- the coverage of word formation by considering the syntactic properties of word constituents -- though it does not always presuppose that affixation can take place in the syntax). If some kinds of affixation occur in the syntax whenever they fulfill the conditions established in it, a new kind of lexicon results, in so far as its rapport with the syntax is dialectic. On the other hand, some word patterns which

have been considered lexical (i.e. given in the lexicon, because of not conforming to the Projection Principle) might in fact be learned, rather than result from linguistic competence. That is, a range of "derivational" processes can be accounted for in the syntax; on the other hand, it is possible that "derivational" relations so far considered to belong to the grammar are not part of a native speaker's competence (though part of his/her knowledge). As a result, word formation in the lexicon might be considerably reduced.

Finally, if, despite its empirical coverage, the attempt to account for word formation in the syntax is considered ill-founded, we will be faced with a problem for the syntactic theory itself, given that with very few extensions it is capable of handling word formation. I have tried to show, however, that a syntactic rendering of word formation (Catalan compounding in particular) is not only possible, but also necessary to give this theory of grammar the generality it allows for.

References

- Allen, M. (1978). *Morphological Investigations*. Doctoral dissertation, University of Connecticut.
- Allwood, J., L-G. Andersson, and O. Dahl. (1977). *Logic in Linguistics*. Cambridge: Cambridge University Press.
- Anderson, J. (1971). *The Grammar of Case*. Cambridge: Cambridge University Press.
- Anderson, S. (1977). Comments on a paper by Wasow. In Culicover, P., T. Wasow, and A. Akmajian (Eds.), *Formal Syntax*. New York: Academic Press.
- Anderson, S. (1982). Where's morphology? *Linguistic Inquiry*, 13, 571-612.
- Aoun, J., and D. Sportiche. (1983). On the formal theory of government. *The Linguistic Review*, 2, 211-236.
- Argente, J. (1976). "Li vaig veure les cuixes". *Els Marges*, 6, 103-110.
- Aronoff, M. (1976). *Word Formation in Generative Grammar*. Cambridge, Mass.: MIT Press.
- Bach, E. (1983). On the relationship between word-grammar and phrase-grammar. *Natural Language and Linguistic Theory*, 1, 65-89.
- Badia i Margarit, A.M. (1962). *Gramàtica catalana*. Madrid: Gredos.
- Baker, M. (1988a). *Incorporation: A Theory of Grammatical Function Changing*. Chicago: University of Chicago Press.
- Baker, M. (1988b). Morphological and syntactic objects: A review of Di Sciullo and Williams' *On the Definition of Word*. In Booij, G., and J. v. Marle (Eds.), *Yearbook of Morphology 1*. Dordrecht: Foris.
- Baker, M. (1988c). Morphology and syntax: An interlocking independence. In Everaert, M., A. Evers, R. Huybregts, and M. Trommelen (Eds.), *Morphology and Modularity*. Dordrecht: Foris.
- Baker, M. (1989). Object sharing and projection in serial verb constructions. *Linguistic Inquiry*, 20, 513-553.
- Bally, C. (1950). *Linguistique générale et linguistique française*. Bern: Francke Verlag.
- Bar-Hillel, Y. (1955). Idioms. In Locke, W., and A. Booth (Eds.), *Machine Translation of Languages. Fourteen Essays*. Cambridge, Mass.: MIT Press, and New York: Wiley and Sons.
- Bartra, A. (1985). *Qüestions de la sintaxi d'ordre en català*. Doctoral dissertation, Universitat Autònoma de Barcelona.
- Bauer, L. (1978). *The Grammar of Nominal Compounding with Special Reference to Danish, English and French*. Odense: Odense Universitets Publikationskonto.
- Bauer, L. (1990). Be-heading the word. *Journal of Linguistics*, 26, 1-32.
- Benveniste, E. (1974a). Formes nouvelles de la composition nominale. In *Problèmes de linguistique générale*, Vol. II. Paris: Gallimard.
- Benveniste, E. (1974b). Fondements syntaxiques de la composition nominale. In

Problèmes de linguistique générale, Vol. II. Paris: Gallimard.

- Berwick, R. (1985). *The Acquisition of Syntactic Knowledge*. Cambridge, Mass.: MIT Press.
- Blasco, E. (1984). *Grammatica storica del catalano e dei suoi dialetti con speciale riguardo all'algherese*. Tübingen: Gunter Narr Verlag.
- Bloomfield, L. (1933). *Language*. New York: Holt, Rinehart and Winston.
- Boase-Beier, J., and J. Toman. (1987). On theta-role assignment in German compounds. In *Neuere Forschungen zur Wortbildung und Historiographie der Linguistik*. Tübingen: Gunter Narr Verlag.
- Bonet, S., and J. Solà. (1986). *Sintaxi generativa catalana*. Barcelona: Enciclopèdia Catalana.
- Bonet, E. (1989). Postverbal subjects in Catalan. Ms., Massachusetts Institute of Technology.
- Borer, H. (1984). The Projection Principle and the rules of morphology. *North Eastern Linguistic Society*, 14, 16-33.
- Borer, H. (1986). Introduction. In Borer, H. (Ed.), *The Syntax of Pronominal Clitics*. New York: Academic Press.
- Borer, H. (1988). On the morphological parallelism between compounds and constructs. In Booij, G., and J. v. Marle (Eds.), *Yearbook of Morphology 1*. Dordrecht: Foris.
- Borer, H. (1990). V+ing: It walks like an adjective, it talks like an adjective. *Linguistic Inquiry*, 21, 95-103.
- Botha, R. (1968). *The Function of the Lexicon in Transformational Generative Grammar*. The Hague: Mouton.
- Botha, R. (1981). A base rule theory of Afrikaans synthetic compounding. In Moortgat, M., H. v. d. Hulst, and T. Hoekstra (Eds.), *The Scope of Lexical Rules*. Dordrecht: Foris.
- Bourdieu, P., J. Passeron, and J. Chamboredon. (1968). *Le métier de sociologue*. Paris: Mouton-Bordas.
- Burton-Roberts, N. (1975). Nominal apposition. *Foundations of Language*, 13, 391-419.
- Burzio, L. (1986). *Italian Syntax*. Dordrecht: Reidel.
- Cabré, T. (1988). La prefixació en català. In Stacsek, J. (Ed.), *On Spanish, Portuguese, and Catalan Linguistics*. Washington D.C.: Georgetown University Press.
- Cabré, T., and G. Rigau. (1986). *Lexicologia i semàntica*. Barcelona: Enciclopèdia Catalana.
- Cann, R. (1987). Splitting heads. Paper presented to the Spring Meeting of the Linguistic Association of Great Britain, University of Bradford.
- Cann, R., and M. Tait. (1990). Rethinking rule R. Paper presented to the Spring Meeting of the Linguistic Association of Great Britain, University of Cambridge.
- Carrier-Duncan, J. (1985). Linking of thematic roles in derivational word formation.

Linguistic Inquiry, 16, 1-34.

- Chomsky, N. (1970). Remarks on nominalization. In Jacobs, A., and P. Rosenbaum (Eds.), *Readings in English Transformational Grammar*. Waltham, Mass.: Ginn and Co.
- Chomsky, N. (1981). *Lectures on Government and Binding*. Dordrecht: Foris.
- Chomsky, N. (1982). *The Generative Enterprise. A Discussion with Riny Huybregts and Henk van Riemsdijk*. Dordrecht: Foris.
- Chomsky, N. (1986a). *Knowledge of Language; Its Nature, Origin and Use*. New York: Praeger.
- Chomsky, N. (1986b). *Barriers*. Cambridge, Mass.: MIT Press.
- Chomsky, N. (1988). Some notes on economy of derivation and representation. Ms., Massachusetts Institute of Technology.
- Chomsky, N., and M. Halle. (1968). *The Sound Pattern of English*. New York: Harper and Row.
- Corbett, G. (1990). Gender resolution. Ms., University of Surrey.
- Corbett, G. (forthcoming). *Gender*. Cambridge: Cambridge University Press.
- Corbin, D. (1987). *Morphologie dérivationnelle et structuration du lexique*. Tübingen: Niemeyer.
- Coromines, J. (1982). *Diccionari etimològic i complementari de la llengua catalana, Vol. III*. Barcelona: Curial and "La Caixa".
- Coseriu, E. (1978). *Gramática, semántica, universales*. Madrid: Gredos.
- Coseriu, E. (1981). *Lecciones de lingüística general*. Madrid: Gredos.
- Cutler, A. (1982). Idioms: The colder the older. *Linguistic Inquiry*, 13, 317-320.
- Darmesteter, A. (1875). *Traité de la formation des mots composés dans la langue française*. Paris: A. Franck.
- Di Sciullo, A.M. (1981). Sur les composés verbaux du français. *Revue de l'Association Québécoise de Linguistique*, 1, 45-59.
- Di Sciullo, A.M., and E. Williams. (1987). *On the Definition of Word*. Cambridge, Mass.: MIT Press.
- Dowty, D. (1979). *Word Meaning and Montague Grammar*. Dordrecht: Reidel.
- Dowty, D. (1989). On the semantic content of the notion "thematic role". In Chierchia, G., B. Partee, and R. Turner (Eds.), *Properties, Types and Meaning, Vol. II*. Dordrecht: Kluwer.
- Dowty, D., R. Wall, and S. Peters. (1981). *Introduction to Montague Semantics*. Dordrecht: Reidel.
- Duarte, C., and A. Alsina. (1986). *Gramàtica històrica del català, Vol. III*. Barcelona: Curial.
- Espinal, M.T. (1990a). Tipologia dels adverbis: El cas dels anomenats adverbis oracionals. In *Miscel·lània Joan Bastardas*. Abadia de Montserrat: Publicacions de l'Abadia de Montserrat.

- Espinal, M.T. (1990b). Nota sobre una tipologia dels adverbis en -MENT. In *Miscel·lània Joan Fuster*. València: Universitat de València, and Abadia de Montserrat: Publicacions de l'Abadia de Montserrat.
- Fabb, N. (1984). *Syntactic Affixation*. Doctoral dissertation, Massachusetts Institute of Technology.
- Fabra, P. (1912). *Gramàtica de la llengua catalana*. Barcelona: L'Avenç.
- Fabra, P. (1932). *Diccionari general de la llengua catalana*. Barcelona: Llibreria Catalònia.
- Fabra, P. (1956). *Gramàtica catalana*. Barcelona: Teide.
- Fabra, P. (1983). *Diccionari manual de la llengua catalana*. Barcelona: Edhasa.
- Fanselow, G. (1988). "Word syntax" and semantic principles. In Booij, G., and J. v. Marle (Eds.), *Yearbook of Morphology 1*. Dordrecht: Foris.
- Ferrater, G. (1981a). Esquelets a l'armari. In Ferraté, J. (Ed.), *Sobre el llenguatge*. Barcelona: Edicions dels Quaderns Crema.
- Ferrater, G. (1981b). La composició nominal. In Ferraté, J. (Ed.), *Sobre el llenguatge*. Barcelona: Edicions dels Quaderns Crema.
- Fillmore, C. (1968). The case for case. In Bach, E., and R. Harms (Eds.), *Universals in Linguistic Theory*. New York: Holt, Rinehart and Winston.
- Foley, W., and R. Van Valin. (1984). *Functional Syntax and Universal Grammar*. Cambridge: Cambridge University Press.
- Gavarró, A. (1988). Catalan compounds and argument structure. *Edinburgh Linguistics Work in Progress*, 21, 20-25.
- Gavarró, A. (1989). On g-Case and word formation. *Edinburgh Linguistics Work in Progress*, 22, 24-29.
- Gavarró, A. (to appear). On idiosyncrasy. *Edinburgh Linguistics Work in Progress*, 23.
- Gavarró, A. (forthcoming). A reply to Scalise on "the notion of head in morphology". In Booij, G., and J. v. Marle (Eds.), *Yearbook of Morphology 3*. Dordrecht: Foris.
- Gazdar, G., E. Klein, G. Pullum, and I. Sag. (1985). *Generalized Phrase Structure Grammar*. Oxford: Basil Blackwell.
- Gràcia, L. (1986). *La teoria temàtica*. Doctoral dissertation, Universitat Autònoma de Barcelona.
- Grimshaw, J. (1981). Form, function and the language acquisition device. In Baker, C., and J. McCarthy (Eds.), *The Logical Problem of Language Acquisition*. Cambridge, Mass.: MIT Press.
- Grossmann, M. (1986). Anàlisi dels compostos catalans del tipus *somiatruïtes*. In *Estudis de llengua i literatura catalana XII*. Abadia de Montserrat: Publicacions de l'Abadia de Montserrat.
- Gruber, J. (1976). *Lexical Studies in Syntax and Semantics*. Amsterdam: North-Holland.
- Guéron, J. (1984). Inalienable possession, PRO-inclusion and lexical chains. In Guéron, J., H. Obenauer, and J.Y. Pollock (Eds.), *Grammatical Representation*. Dordrecht: Foris.

- Halle, M. (1973). Prolegomena to a theory of word formation. *Linguistic Inquiry*, 4, 3-16.
- Halle, M., and K. Mohanan. (1985). Segmental phonology of modern English. *Linguistic Inquiry*, 16, 57-116.
- Higginbotham, J. (1985). On semantics. *Linguistic Inquiry*, 16, 547-593.
- Hoeksema, J. (1985). *Categorial Morphology*. New York: Garland Publishing.
- Hoeksema, J. (1988). Head-types in morphosyntax. In Booij, G. and J. v. Marle (Eds.), *Yearbook of Morphology 1*. Dordrecht: Foris.
- Hudson, R. (1987). Zwicky on heads. *Journal of Linguistics*, 23, 109-132.
- Jackendoff, R. (1972). *Semantic Interpretation in Generative Grammar*. Cambridge, Mass.: MIT Press.
- Jackendoff, R. (1975). Morphological and semantic regularities in the lexicon. *Language*, 51, 639-671.
- Jackendoff, R. (1976). Towards an explanatory semantic representation. *Linguistic Inquiry*, 7, 89-150.
- Jaeggli, O. (1982). *Topics in Romance Syntax*. Dordrecht: Foris.
- Jayaseelan, K. (1988). Complex predicates and θ -theory. In Wilkins, W. (Ed.), *Thematic Relations*. San Diego: Academic Press.
- Jensen, J., and M. Stong-Jensen. (1984). Morphology is in the lexicon! *Linguistic Inquiry*, 15, 474-498.
- Kageyama, T. (1989). The place of morphology in the grammar: Verb-verb compounds in Japanese. In Booij, G., and J. v. Marle (Eds.), *Yearbook of Morphology 2*. Dordrecht: Foris.
- Kaye, J., J. Lowenstam, and J.R. Vergnaud. (1985). The internal structure of phonological elements: A theory of charm and government. In Ewen, C., and J. Anderson (Eds.), *Phonology Yearbook 2*. Cambridge: Cambridge University Press.
- Kayne, R. (1975). *French Syntax: The Transformational Cycle*. Cambridge, Mass.: MIT Press.
- Kayne, R. (1982). Predicates and arguments, verbs and nouns. *GLOW Newsletter*, 8, 24.
- Kayne, R. (1984). *Connectedness and Binary Branching*. Dordrecht: Foris.
- Kayne, R. (1989). Null subjects and clitic climbing. In Jaeggli, O., and K. Safir (Eds.), *The Null Subject Parameter*. Dordrecht: Kluwer.
- Kiparsky, P. (1982). *Lexical Phonology and Morphology*. Ms., Massachusetts Institute of Technology.
- Kiparsky, P. (1985). Some consequences of lexical phonology. In Ewen, C., and J. Anderson (Eds.), *Phonology Yearbook 2*. Cambridge: Cambridge University Press.
- Lapointe, S. (1981). General and restricted agreement phenomena. In Moortgat, M., H. v. d. Hulst, T. Hoekstra (Eds.), *The Scope of Lexical Rules*. Dordrecht: Foris.

- Larson, R. (1985). Bare-NP adverbs. *Linguistic Inquiry*, 16, 595-621.
- Larson, R. (1987). "Missing prepositions" and the analysis of English free relative clauses. *Linguistic Inquiry*, 18, 239-266.
- Leech, G. (1974). *Semantics*. Harmondsworth: Penguin.
- Lees, R. (1960). *The Grammar of English Nominalizations*. The Hague: Mouton.
- Levin, B., and M. Rappaport. (1986). The formation of adjectival passives. *Linguistic Inquiry*, 17, 623-661.
- Lieber, R. (1980). *On the Organization of the Lexicon*. Doctoral dissertation, Massachusetts Institute of Technology.
- Lieber, R. (1981). Morphological conversion within a restrictive theory of the lexicon. In Moortgat, M., H. v. d. Hulst, and T. Hoekstra (Eds.), *The Scope of Lexical Rules*. Dordrecht: Foris.
- Lieber, R. (1983). Argument linking and compounds in English. *Linguistic Inquiry*, 14, 251-285.
- Lieber, R. (1984). Grammatical rules and sublexical elements. In Testen, D., V. Mishra, and J. Drogo (Eds.), *Papers from the Parasession on Lexical Semantics*. Chicago Linguistic Society.
- Lieber, R. (1989). On percolation. In Booij, G., and J. v. Marle (Eds.), *Yearbook of Morphology 2*. Dordrecht: Foris.
- Manzini, M. R. (1986). On Italian *si*. In Borer, H. (Ed.), *The Syntax of Pronominal Clitics*. New York: Academic Press.
- Marvà, J. (1934). *Curs superior de gramàtica catalana*. Barcelona: Barcino.
- Mascaró, J. (1975). Perceptual level of nonmain stress in Catalan. Ms., Massachusetts Institute of Technology.
- Mascaró, J. (1983). *La fonologia catalana i el cicle fonològic*. Bellaterra: Publicacions del Departament de Filologia Hispànica de la Universitat Autònoma de Barcelona.
- Mascaró, J. (1986). *Morfologia*. Barcelona: Enciclopèdia Catalana.
- Matthews, P. (1981). *Syntax*. Cambridge: Cambridge University Press.
- McCarthy, J. (1981). A prosodic theory of non-concatenative morphology. *Linguistic Inquiry*, 12, 373-418.
- Mithun, M. (1984). The evolution of noun incorporation. *Language*, 60, 847-95.
- Moll, F. de B. (1975). *Gramàtica catalana referida especialment a les Illes Balears*. Palma de Mallorca: Editorial Moll.
- Montague, R. (1974). The proper treatment of quantification in ordinary English. In Thomason, R. (Ed.), *Formal Philosophy. Selected Papers of Richard Montague*. New Haven: Yale University Press.
- Namiki, T. (1982). The notion "head of a word" and core and periphery word formation: Interactions between affixation and subcategorization. *Studies in English Linguistics*, 10, 21-41.
- Negrão, E. (1990). Changes in the INFL system of Brazilian Portuguese and the pro-

- drop parameter. Paper presented to the Parametric Variation Workshop, Center for Cognitive Science, University of Edinburgh.
- Newmeyer, F. (1986). *Linguistic Theory in America*. San Diego: Academic Press.
- Nichols, J. (1986). Head-marking and dependent-marking grammar. *Language*, 62, 56-119.
- Pesetsky, D. (1982). *Paths and Categories*. Doctoral dissertation, Massachusetts Institute of Technology.
- Pesetsky, D. (1985). Morphology and logical form. *Linguistic Inquiry*, 16, 193-246.
- Postal, P. (1962). *Some Syntactic Rules of Mohawk*. Doctoral dissertation, Yale University.
- Postal, P. (1969). Anaphoric islands. In *Papers from the Fifth Regional Meeting of the Chicago Linguistic Society*. Chicago Linguistic Society.
- Radford, A. (1984). *Transformational Syntax*. Cambridge: Cambridge University Press.
- Rappaport, M., M. Laughren and B. Levin. (1987). Levels of lexical representation. *Lexicon Project Working Papers*, Vol. 20.
- Rappaport, M., and B. Levin. (1988). What to do with θ -roles. In Wilkins, W. (Ed.), *Thematic Relations*. San Diego: Academic Press.
- van Riemsdijk, H., and E. Williams. (1986). *Introduction to the Theory of Grammar*. Cambridge, Mass.: MIT Press.
- Rigau, G. (1990). The semantic nature of some Romance prepositions. In Mascaró, J., and M. Nespor (Eds.), *Grammar in Progress*. Dordrecht: Foris.
- Rizzi, L. (1986). Null objects in Italian and the theory of *pro*. *Linguistic Inquiry*, 17, 501-557.
- Roberts, I. (1985). Agreement parameters and the development of English modal auxiliaries. *Natural Language and Linguistic Theory*, 3, 21-58.
- Roeper, T. (1988). Compound syntax and head movement. In Booij, G., and J. v. Marle (Eds.), *Yearbook of Morphology 1*. Dordrecht: Foris.
- Roeper, T., and M. Siegel. (1978). A lexical transformation for verbal compounds. *Linguistic Inquiry*, 9, 199-260.
- Rothstein, S. (1983). *The Syntactic Forms of Predication*. Doctoral dissertation, Massachusetts Institute of Technology.
- Sadock, J. (1983). The necessary overlapping of grammatical components. In Richardson, J., M. Marks, and A. Chuckerman (Eds.), *Papers from the Parasession on the Interplay of Phonology, Morphology and Syntax*. Chicago Linguistic Society.
- Sadock, J. (1985). Autolexical syntax: A proposal for the treatment of noun incorporation and similar phenomena. *Natural Language and Linguistic Theory*, 3, 379-439.
- Sadock, J. (1986). Some notes on noun incorporation. *Language*, 62, 19-31.
- de Saussure, F. (1916). *Cours de linguistique générale*. Paris: Payot.

- Scalise, S. (1984). *Generative Morphology*. Dordrecht: Foris.
- Selkirk, E. (1981). English compounding and the theory of word structure. In Moortgat, M., H. v. d. Hulst, and T. Hoekstra (Eds.), *The Scope of Lexical Rules*. Dordrecht: Foris.
- Selkirk, E. (1982). *The Syntax of Words*. Cambridge, Mass.: MIT Press.
- Shibatani, M., and K. Kageyama. (1988). Word formation in a modular theory of grammar. *Language*, 64, 451-484.
- Sproat, R. (1985). *On Deriving the Lexicon*. Doctoral dissertation, Massachusetts Institute of Technology.
- Sproat, R., and G. Ward. (1987). Pragmatic considerations in anaphoric island phenomena. In *Proceedings of the 23rd Regional Meeting*. Chicago Linguistic Society.
- Stowell, T. (1981). *Origins of Phrase Structure*. Doctoral dissertation, Massachusetts Institute of Technology.
- Sugioka, Y. (1984). *Interaction of Derivational Morphology in Japanese and English*. Doctoral dissertation, University of Chicago.
- Szabolcsi, A. (1987). Functional categories in the noun phrase. In Kenesei, I. (Ed.), *Approaches to Hungarian. Theories and Analyses*. Szeged: Jate.
- Tait, M., and R. Cann. (1989). On licensing empty heads. Ms., University of Edinburgh.
- Travis, L. (1989). Parameters of phrase structure. In Baltin, M., and A. Kroch (Eds.), *Alternative Conceptions of Phrase Structure*. Chicago: University of Chicago Press.
- Trommelen, M., and W. Zonneveld. (1986). Dutch morphology: Evidence for the right-hand head rule. *Linguistic Inquiry*, 17, 147-169.
- Vallduví, E. (1988). Functional load, prosody and syntax: Left-detachment in Catalan and Spanish. In *Papers from the 24th Annual Regional Meeting*. Chicago Linguistic Society.
- Wheeler, M. (1977). Problemes lèxics per al generativista: Els mots catalans a doble arrel. In *Actes del IV Col·loqui Internacional de Llengua i Literatura Catalanes*. Abadia de Montserrat: Publicacions de l'Abadia de Montserrat.
- Wheeler, M. (1979). *Phonology of Catalan*. Oxford: Basil Blackwell.
- Williams, E. (1980). Predication. *Linguistic Inquiry*, 11, 203-238.
- Williams, E. (1981a). Argument structure and morphology. *Linguistic Review*, 1, 81-114.
- Williams, E. (1981b). On the notions "lexically related" and "head of a word". *Linguistic Inquiry*, 12, 245-274.
- Williams, E. (1989). Maximal projections in words and phrases. In Baltin, M., and A. Kroch (Eds.), *Alternative Conceptions of Phrase Structure*. Chicago: University of Chicago Press.
- Williams, M. (1976). *A Grammar of Tuscarora*. New York: Garland.
- Zubizarreta, M. L. (1985). The relation between morphophonology and morphosyntax:

- The case of Romance causatives. *Linguistic Inquiry*, 16, 247-289.
- Zwanenburg, W. (1989). Review of Danielle Corbin, *Morphologie dérivationnelle et structuration du lexique*. In Booij, G., and J. v. Marle (Eds.), *Morphology Yearbook*, 2. Dordrecht: Foris.
- Zwicky, A. (1985). Heads. *Journal of Linguistics*, 21, 1-30.